

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

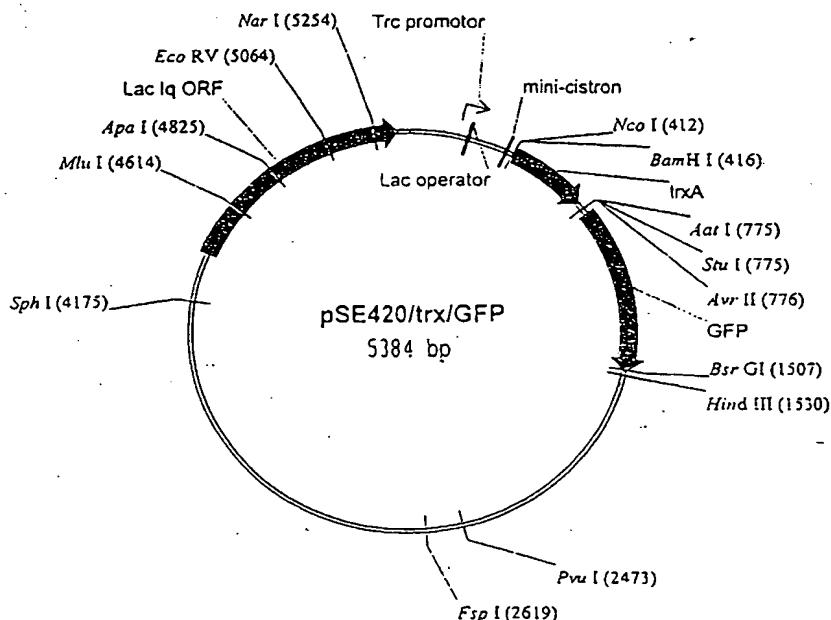


61

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ :	A1	(11) International Publication Number: WO 00/49417
G01N 33/68		(43) International Publication Date: 24 August 2000 (24.08.00)
(21) International Application Number:	PCT/US00/04089	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LÜ, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date:	17 February 2000 (17.02.00)	
(30) Priority Data:		
09/251,364	17 February 1999 (17.02.99)	US
09/350,419	8 July 1999 (08.07.99)	US
(71) Applicant:	ARCARIS, INC. [US/US]; Suite 300, 615 Arapeen Drive, Salt Lake City, UT 84108 (US).	
(72) Inventor:	KAMB, Carl, Alexander; 686 Donner Hill Circle, Salt Lake City, UT 84108 (US).	
(74) Agent:	RIN-LAURES, Li-Hsien; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606-6402 (US).	

(54) Title: METHODS FOR SUBSTRATE-LIGAND INTERACTION SCREENING



(57) Abstract

Provided by the present invention are novel methods of detecting substrate-ligand interactions, and more specifically relates to methods for detecting and characterizing polypeptide-ligand interactions. By practice of this invention, protein interaction maps may be generated for humans or for other organisms.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

1 METHODS FOR SUBSTRATE-LIGAND INTERACTION SCREENING

2

3

4

RELATED APPLICATIONS

5

6

7

8

9

This application is a continuation in part of US application serial nos. 09/251,364 and 09/350419 of K. A. Kamb, entitled "Methods For Substrate-Ligand Interaction Screening," and claims priority therefrom. The disclosures of the priority applications are incorporated by reference in their entirety herein.

10

FIELD OF THE INVENTION

11

12

13

14

15

The present invention relates generally to novel methods of screening for, detecting, identifying and quantifying substrate-ligand interactions, and more specifically relates to novel methods for achieving these ends for protein-ligand interactions, and more specifically protein-protein interactions. The inventive method is suitable for screening large or very large libraries, and for generating protein interaction maps.

16

17

BACKGROUND OF THE INVENTION

18

19

20

21

22

23

24

25

26

27

Many physiological functions in mammals and other organisms are mediated through interactions of cellular proteins with a variety of endogenous ligands, including for example other proteins, glycoproteins, polypeptides, hormones or other small molecules. Because of the importance of these endogenous protein-ligand interactions, pharmaceutical companies often seek to modify or disrupt physiological pathways by providing exogenous molecules that interact with those endogenous proteins. In some cases, researchers may target particular, previously characterized proteins, and screen for molecules that interact with that protein. But in the vast majority of cases, researchers lack the initial insight into a given physiological pathway, and must first identify the native proteins involved in that pathway before achieving the ability to modify the physiological effects of that pathway.

28

29

30

31

32

33

34

35

While much is now known about the genome of humans and other organisms, researchers have yet to close the link in many instances between DNA sequence information and physiological function. In order to do so efficiently, it is desirable to first identify key native proteins that are related to specific physiological functions, and then to relate those proteins to the DNA sequences encoding them. Once such key proteins are identified, then researchers may identify ligands (proteinaceous or otherwise) that interact with these proteins, and in turn relate these targeted protein-ligand interactions to physiological changes. But to date, the methods used in the art for evaluating protein-ligand interactions have not

1 provided a simple, efficient method of identifying the key native proteins (and screening for
2 ligands that interact with them). Nor has the art provided an efficient high-throughput
3 screening method that allows researchers to broadly catalogue, e.g., all endogenous protein-
4 protein interactions, before turning to the related questions of physiological function and
5 targeted drug development.

6 Researchers are particularly hampered in their ability to comprehensively catalogue
7 endogenous protein-protein interactions in a human or other organism by the sheer magnitude
8 of endogenous proteins that must be evaluated. For example, some $10^5 - 10^6$ proteins are
9 believed to be encoded by the human genome. To begin by evaluating the interaction of each
10 of those proteins with each other encoded protein thus requires evaluating $10^6 \times 10^6$ protein-
11 protein interactions, or 10^{12} total interactions. Such a large-scale evaluation is problematic,
12 because it involves evaluating a matrix of all possible combinations; thus the number of
13 interactions scales as the square of the number of proteins to be evaluated (termed more
14 generally herein, the "n x n" problem). Current methodologies simply cannot evaluate such
15 vast numbers of protein interactions in a time- and cost-efficient manner. The inability of
16 current methodologies to provide rapid, quantitative high-throughput screening is particularly
17 acute if comparative information regarding protein interactions in different cell types or cell
18 states is desired.

19 The limitations of current methodologies can be seen by considering current
20 technologies for mapping protein interactions. For example, one typical approach to probing
21 protein-ligand interactions involves an *in vivo*, quasi-genetic approach known as the yeast
22 two-hybrid assay. This approach suffers from the drawbacks of (i) limitation to probing
23 protein-protein interactions, (ii) lack of speed, (iii) prevalence of false-positive and false-
24 negative results, (iv) lack of quantitative information (e.g., binding affinities between specific
25 protein pairs). These drawbacks remain a substantial obstacle to utilizing yeast two-hybrid
26 technology to screen for interactions, notwithstanding recent advances in, e.g., automation of
27 the two-hybrid technology.

28 Phage display techniques have been used to select proteins that bind to a particular,
29 pre-selected ligand. Such methodologies again are essentially *in vivo*, as the proteins that are
30 borne by the phages are isolated and identified only after the intermediate steps of culturing
31 the phage in *E. coli*, plating the bacteria and isolating phage from phage-generated plaques or
32 cultures. These intermediate steps are necessary because the phage must be generated in cells
33 and cannot be created without cells. In addition, phage must be bound, eluted, and re-grown

1 in cells prior to analysis. Thus, the technique is not well suited to screening applications such
2 as generating protein interaction maps. Nor is the technique amenable to high throughput
3 applications. Moreover, the technique does not provide quantitative information.

4 Alternatively, researchers have utilized limited-throughput screening techniques to
5 evaluate the binding of ligands to a particular substrate. For example, a selected
6 proteinaceous substrate, or small number of such substrates, have been immobilized by a
7 variety of means for exposure to a select pool of ligands. E.g., U.S. No. 5,635,182; US No.
8 5,776,696; US No. 5,498,530; Major, E.S., "Challenges of high throughput screening against
9 cell surface receptors," *J. Recept. Signal Transduct. Res.* 15(1-4):595-607 (1995). But such
10 methodologies are not amenable to screening, e.g., large or very large populations, for
11 generating protein interaction maps, and/or for screening previously uncharacterized
12 substrates -- i.e., the techniques do not adequately address the " $n \times n$ " problem generated by
13 large-scale screening efforts.

14 More generally, other researchers have utilized various solid-state screening
15 techniques to evaluate interactions of different moieties. For example, assays exist that
16 immobilize known antigens or antibodies on beads or other such solid supports. E.g., Roque
17 et al., *Acta Histochem.* 98(4):441-451 (Nov. 1996). Two or three-dimensional matrices
18 tagged with nucleic acids have been utilized to screen for DNA-binding moieties. Other
19 researchers have utilized "lawn assays" that detect protein interactions utilizing diffusion of a
20 ligand through a colloidal matrix. However, none of these techniques addresses the " $n \times n$ "
21 problem, and thus none provides rapid, quantitative and/or large-scale evaluation of
22 substrate-ligand interaction, or more specifically, protein-protein interactions.

23 Thus, the need remains for a flexible, efficient, quantitative methodology for
24 evaluating substrate-ligand interactions generally, and protein-protein interactions in
25 particular. The present invention meets such needs.

SUMMARY OF THE INVENTION

26 The present invention provides methods for detecting substrate-ligand interactions,
27 more particularly polypeptide-ligand interactions or polypeptide-polypeptide interactions.
28 The polypeptides may be individual polypeptides, or may alternatively be library
29 polypeptides, including those of large or very large libraries and/or of native, endogenous
30 polypeptides. The methods utilize randomizable ligand-bearing supports bearing unique tags,
31 and may optionally use location-determinable supports. In some embodiments, a magnetic
32 support may be used to adhere to either the substrate or the ligand, and magnetic culling of
33 support may be used to adhere to either the substrate or the ligand, and magnetic culling of

1 bead aggregates that result from substrate-ligand complexes provides for an enrichment step.
2 Interacting pairs are identified by correlating (i) location information and (ii) identity
3 information provided by each unique tag. The location information may be derived from
4 correlating back to a unique location, or alternatively by evaluating the origination of
5 location-determinable supports. The unique tags may use a variety of techniques, including
6 fluorescent bar codes, to encode ligand identity information. By such methods, protein
7 interaction maps for, e.g., the human organism, may be generated.

8 The invention further provides methods for identifying and quantifying such
9 interactions. In some embodiments, the interacting substrate-ligand pairs may be detected
10 with antibodies, for example fluorescent antibodies, and the interactions quantified via a
11 FACS machine or CCD camera.

BRIEF DESCRIPTION OF THE DRAWINGS

12 **FIGURE 1** is a map of plasmid vector pSE420/trx/GFP.
13

14 **FIGURE 2** is a map of plasmid vector pSE420/biotrx/GFP/BirA.
15

16 **FIGURE 3** is a map of plasmid vector pSE420/Caltrx/GFP.
17

18 **FIGURE 4** is a map of plasmid vector pSE420/DHFR/GFP.
19

20 **FIGURE 5** is a map of plasmid vector pLex biotrx GFP LbirA.
21

22 **FIGURE 6** depicts a bead that has been derivatized for crosslinking with a
23 methotrexate as an adhesion moiety and SANPAH as a photoactivatable crosslinker.
24

25 **FIGURE 7** is a FACS histogram demonstrating the crosslinking of interacting
26 proteins. Peak A is streptavidin coated particles reacted with BL21 lysate and FITC-
27 calmodulin conjugate. Peak B is streptavidin coated particles reacted with a lysate having a
28 biotin-thioredoxin-CBP fusion protein, which is then exposed to the FITC-calmodulin
29 conjugate in the presence of calcium chelator EGTA. Peak C is streptavidin coated particles
30 reacted with a lysate having a biotin-thioredoxin-CBP fusion protein, which is then exposed
31 to a FITC-calmodulin conjugate. Peak D is streptavidin coated particles reacted with a lysate
32 having a biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate and a protein
33 crosslinking agent. Peak E is streptavidin coated particles reacted with a lysate having a
biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate, protein crosslinking
agent and then EGTA.

34 **FIGURE 8** depicts the enrichment of biotin-coated fluorescent beads from a mixture
35 of fluorescent beads coated only with Bovine Serum Albumin (BSA), using streptavidin-

1 coated magnetic beads. The streptavidin and the biotin interact, and subsequently the
2 aggregates are segregated from the BSA-coated beads with a magnet.

3 FIGURE 9 depicts the enrichment of beads coated with an SV40 large T antigen
4 conjugate from a mixture of fluorescent beads coated only with BSA, using magnetic beads
5 coated with an anti-SV40 large T antigen antibody conjugate. The antigen and antibody
6 interact, and subsequently the aggregates are segregated from the BSA-coated beads with a
7 magnet.

8

9 **DETAILED DESCRIPTION OF THE INVENTION**

10 The methodologies of this invention provide rapid, efficient, quantitative substrate-
11 ligand interaction screens. The invention differs from prior approaches in that it does not rely
12 on yeast-two hybrid technology or other such *in vivo* techniques, but instead provides a high
13 throughput *in vitro* screening methodology. While the inventive methods do provide rapid,
14 quantitative screening of individual polypeptides or other substrates against a selected ligand
15 pool, the techniques provide for scale-up for screening small (on the order of 1×10^2)
16 substrate populations, and advantageously may be used to screen large (on the order of 10^3 or
17 10^4) or even very large (10^5 , 10^6 or even 10^7) populations. This is so because the inventive
18 use of both location information and unique tags to identify substrate-ligand pairs renders the
19 technique suitable for screening previously uncharacterized polypeptides or other substrates
20 *en masse*, rather than relying upon the pre-selection of a known substrate or small number of
21 substrates and thereby screening in a "1xn" manner rather than an "n x n" manner.

22 More specifically, the invention provides its quantitative, high throughput
23 polypeptide/ligand screening capabilities by cross-indexing (i) polypeptide (or other
24 substrate) identity information derived from the characteristic, unique location from which
25 one particular polypeptide (or other substrate) is derived, and (ii) ligand identity information
26 derived from its associated randomizable support, which bears a unique tag that correlates to
27 the identity of that ligand. The polypeptide may be an individual polypeptide, or alternatively
28 may be a member of a polypeptide library of various sizes. Non-polypeptide substrates may
29 include, e.g., small organic or inorganic molecules, of either endogenous or synthetic origin.

30 In some embodiments, a unique polypeptide or other such substrate may be adhered to
31 a location-determinable support, which correlates to the unique location from which a
32 particular library polypeptide is derived, prior to exposure to the ligands. In other
33 embodiments the unique polypeptide or substrate remains in a lysate or other such solution,

1 to which the randomizable ligand-bearing supports are added. The supports described herein
2 may be microbeads, or may be a fixed solid support. The unique tag that identifies a
3 particular ligand may be, for example, a fluorescent "bar code" or oligonucleotide tag.

4 The invention encompasses a number of potential substrates, including (i) non-nucleic
5 acid, proteinaceous substrates such as individual polypeptides and library polypeptides, (ii)
6 other non-nucleic acid substrates such as exogenous natural products, exogenous small
7 organic molecules or endogenous non-proteinaceous products, (iii) nucleic acid substrates,
8 and (iv) inorganic substrates. The term "individual polypeptide" refers to an amino acid
9 sequence, for example a protein or protein domain, and also includes further derivatized
10 amino acid sequences, such as, e.g., glycoproteins. The sequence may be that of a native
11 molecule (i.e., endogenous to a given cell), or alternatively may be synthetic. Individual
12 polypeptides are typically identified and characterized in advance of the ligand screening, and
13 are not generated or screened *en masse*. Library polypeptides encompass the same sorts of
14 amino acid sequences, but are encoded by DNA sequences that are generated and screened *en*
15 *masse*, and may be previously unknown or uncharacterized molecules. The libraries may
16 vary in size, and include large or very large libraries. In particular, the library polypeptides
17 may include all or substantially all native protein domains encoded by the human genome, or
18 expressed in the human organism. As termed herein, "ligands" are molecules that are
19 screened to identify those members that interact with the polypeptides or other substrates.
20 Ligands may be proteinaceous moieties such as, e.g., polypeptides or glycoproteins from a
21 variety of sources, or may be other organic or inorganic molecules. The ligands may be
22 endogenous molecules such as hormones, antibodies, receptors, peptides, enzymes, growth
23 factors or cellular adhesion molecules, or may be derivatized or wholly synthetic molecules.
24 Because of the flexibility of the invention, the identity of the ligands need not be known or
25 pre-selected in advance, and may also be large or very large populations.

26 The present invention lends itself to automated high-throughput embodiments, in
27 which microbeads serve as the location-determinable and/or randomizable supports. Such
28 microbeads may be readily dispersed by robotic means to, e.g., 384-well microtiter plates.
29 The polypeptides or other such substrates interact with ligands to form interacting pairs,
30 termed "complexes" herein. When each member of the interacting pair are immobilized on
31 supports, then the two supports are linked via the substrate/ligand complex to form an
32 "aggregate." The aggregates and/or complexes are then sorted and identified. Means for
33 accomplishing this include a CCD camera or a fluorescence-activated cell sorter (FACS).

1 The speed and selectivity of this inventive methodology may be further enhanced by
2 utilizing magnetic attraction to facilitate a solid-state interaction between the polypeptide or
3 other substrate that is bound to a location-determinable support, and the ligand that is bound
4 to the randomizable support. This may be accomplished by utilizing a magnetic material for
5 the support, and then collecting the complexes or aggregates by culling the magnetic supports
6 with a magnetic force, for example by applying a magnetic field to the exterior of the arrays
7 or by inserting a magnetized body such as a pin into each well of the array.

8 Because the methodologies of the present invention are so rapid and efficient,
9 screening is not limited to small, pre-characterized or artificially culled substrate populations,
10 nor does the invention require pre-selection of known ligands of interest. Rather, the
11 invention allows for high throughput cross-screening of large or very large populations --
12 e.g., the entire endogenous protein library of a human organism. Indeed, the methodologies
13 of this invention are particularly well-suited for large-scale screening of some 1×10^6
14 proteins, which is the estimated number of proteins produced in a human being. Thus, the
15 inventive methods and materials answer a long-felt need in the industry for evaluating the
16 interactions of endogenous proteins within a human organism, to form a comprehensive
17 human "protein interaction map." Alternatively, the inventive methodology may be used to
18 screen the selected library polypeptides against other ligand libraries -- for example,
19 endogenous ligand libraries such as a second polypeptide library, endogenous hormones,
20 antibodies, receptors, peptides, enzymes, growth factors or cellular adhesion molecules, or on
21 the other hand exogenous ligands derivatized or wholly synthetic molecules, natural products,
22 synthetic peptides, or synthetic organic or inorganic molecules.

23 Other uses and advantages of this screening methodology will be apparent to those of
24 skill in the art.

25 **Overview of the methodology**

26 The general strategy of the methodology is exemplified as follows. A substrate pool
27 of interest is selected -- for example, a library of all or substantially all native polypeptides
28 expressed by the human organism, or a selection of individual polypeptides of interest. A
29 corresponding set of library polypeptides or individual polypeptides are generated in cells.
30 Single colonies, each of which is expressing one particular polypeptide of interest, are
31 selected and replated in order to generate single-cell clones (i.e., multiple copies of one
32 particular cell, each cell expressing the same individual polypeptide or unique member of the
33 polypeptide library). Each such clone is uniquely located at one particular location of an

1 array -- e.g., each particular well of a given 384 well plate contains a one particular clone.
2 The expression products of each of those clones are then harvested from the cells, for
3 example by generating soluble lysates that correspond to each of the plated clones. Thus,
4 each well corresponds to the soluble lysate of one particular clone, which in turn corresponds
5 to one individual polypeptide or one unique member of a polypeptide library. Alternatively,
6 each member of a non-proteinaceous substrate pool of interest is individually arrayed at a
7 unique location.

8 In the case of proteinaceous substrates, the expression product of each lysate is then
9 either (i) kept segregated in a unique location (e.g., one particular well of a 384 well array);
10 or (ii) exposed to a solid support that is unique to that lysate source, and whose location may
11 be tracked in order to identify the corresponding lysate source to which it was exposed. Such
12 a solid support is termed herein, a "location-determinable support." This location-
13 determinable support may be any solid support that is suitable for adhering a desired
14 polypeptide from a polypeptide-containing lysate, and which can be correlated back to a
15 particular polypeptide source -- e.g., a particular microtiter well in a particular array.
16 Exemplary location-determinable supports include (i) beads that are kept segregated in
17 microtiter wells that are derived from, and thus correspond to, the original lysate-bearing
18 array location; and (ii) a fixed solid support such as a pin or other such probe that is suitable
19 for dipping into one unique location in a lysate-bearing microtiter well. The same strategy
20 may be applied to non-proteinaceous substrates.

21 The ligands to be screened may advantageously may be immobilized on a solid
22 support, although in order to screen a large variety of ligands for interaction with any
23 particular substrate, such solid supports should be "randomizable" -- i.e., in terms of this
24 invention, (i) each such support can be dispersed into a mixture of such supports in a manner
25 that allows for full mixing and resultant random distribution of support constructs in any
26 subsequent aliquot of the mixture, and (ii) each such randomizable support bears with it a
27 corresponding unique identification tag that identifies the associated ligand. Use of such
28 randomizable supports to create a fully integrated set of ligand-bearing supports increases the
29 statistical likelihood that an aliquot taken from the fully integrated ligand set will contain a
30 fully dispersed, representative subset of ligands. Examples of such randomizable supports
31 include microparticles (e.g., small beads) in a variety of materials and sizes. The unique tags
32 may be, for example, fluorescent, oligonucleotide sequence tags, mass tags, radio tags, or any
33 combination thereof.

1 As one exemplary use of the invention, a polypeptide library may be screened against
2 itself to generate a "protein interaction map" -- i.e., an "n x n" matrix of interactions for all or
3 substantially all native polypeptides of a human or other selected organism. By "native
4 polypeptides" is meant polypeptides that are endogenous to a selected organism -- i.e., that
5 are encoded by the organism's genome and which may be expressed by that organism. Native
6 polypeptides include functional subunits or "protein domains" of endogenous proteins. In
7 such embodiments, the polypeptides of interest serve as both substrate and ligand - i.e., each
8 randomizable support is adhered to multiple copies of one member of the polypeptide library,
9 and each unique array location contains multiple copies of one member of the polypeptide
10 library. Once each randomizable support bears its corresponding unique library polypeptide,
11 the supports are pooled into one volume and mixed to form a fully integrated ligand
12 collection -- i.e., the pooled volume represents all ligand species. Next, ligand aliquots are
13 drawn from this fully integrated ligand collection. Each aliquot contains a randomized,
14 representative sampling of the ligands that is statistically likely to contain at least one copy of
15 each species of ligand present in the pooled ligand volume. These ligand aliquots then are
16 presented for interaction with each of the library polypeptides, either by simply adding an
17 aliquot of integrated ligand-bearing supports to each uniquely located library polypeptide
18 lysate within the library array, or by first adhering the library polypeptides in the array to
19 location-determinable supports and then exposing each such set of polypeptide-bearing
20 supports (which bear only one type of polypeptide) to an integrated aliquot of randomizable
21 supports.

22 In another exemplary use a first set of library polypeptides may be screened against a
23 second, independent polypeptide library, composed of, e.g., a separate set of native protein
24 domains, a set of synthetic polypeptides containing, e.g., point mutations, or randomly
25 generated synthetic polypeptide sequences. In such embodiments, the same methodology is
26 applied, but a second, independent expression library is used to generate a second,
27 independent array containing the second, independent polypeptide library.

28 In another exemplary use, a first set of polypeptides may be screened against some
29 other ligand set -- e.g., small organic molecules, natural products, hormones, receptors,
30 antibodies, peptides, enzymes, growth factors, cellular adhesion molecules, combinatorial
31 library components and the like -- that is adhered to the randomizable support and presented
32 to the library polypeptides. In many such instances, a prior cellular expression step to
33 produce the ligands will not be necessary.

1 Whatever the source of the ligands that are adhered to the randomizable supports, the
2 methodology is completed by exposing each uniquely located substrate (either in solution or
3 adhered to its analogous location-determinable support) to an aliquot of ligand-bearing
4 supports. If the ligand bearing support is exposed directly to a substrate, e.g., to a lysate or
5 other such polypeptide-bearing solution, then any interactions will result in formation of a
6 substrate-ligand complex -- e.g., a randomizable support with consecutive layers of adhered
7 ligand and polypeptide. If the substrate is first immobilized on its own support, then any
8 substrate-ligand interaction will adhere the two supports into an aggregate. Such aggregates
9 may be detected and characterized in that form. Alternatively, the aggregates may be re-
10 suspended in a corresponding unique library polypeptide solution to displace the support-
11 linked polypeptide with an unbound form of that polypeptide, or removed by some other
12 procedure.

13 Interactions between substrates and ligands are then detected by fluorescent or other
14 means, for example by use of a fluorescently tagged antibody. Interacting pairs are then
15 culled out in a sorting or detection process, for example via FACS, so that the components of
16 the various complexes may be identified. The identity of the substrate is determined by
17 correlating it to the unique array location from which it was derived (either directly, or via the
18 analogous location-determinable support). If the substrate is proteinaceous, then the DNA
19 encoding the polypeptide produced by the original single-cell clone at that unique location of
20 the library array may then be sequenced or otherwise characterized. The identity of the
21 ligand is determined by evaluating the associated unique identification tag on the
22 randomizable support to which that ligand is bound. If the ligands are also polypeptides that
23 have been uniquely arrayed, the unique identification tag can be further correlated back to a
24 single clone in its corresponding array location.

25 The screening methods of the present invention can be adapted in a number of ways
26 apparent to those of skill in the art to displacement screening. In one non-limiting
27 embodiment, the substrate-ligand pairs are first formed, and are adhered to a solid support.
28 Subsequently, these pairs are exposed to a secondary ligand. If the secondary ligand is
29 capable of adhering to the substrate, then in many cases it will displace the first ligand. The
30 substrate-secondary ligand pair can then be manipulate, enriched and analyzed according to
31 the method of the invention. The secondary ligand may be a proteinaceous moiety such as,
32 e.g., a polypeptide or glycoprotein from a variety of sources, or may be some other organic or
33 inorganic molecule. The secondary ligand also may be an endogenous molecule such as a

1 hormone, antibody, receptor, peptide, enzyme, growth factor or cellular adhesion molecule,
2 or may be a derivatized or wholly synthetic molecule. In particularly preferred embodiments
3 of displacement screening, the secondary ligand is a small organic molecule.

4

5 **Generation and expression of polypeptide fusion libraries**

6 If the substrate of interest is proteinaceous, then an expression library may be
7 generated first. The overall goal of this step is to generate a selection of desired individual
8 polypeptides or library polypeptides that are suitable as either substrate or ligand (or both),
9 for rapid, efficient ligand interaction screening. Once a desired pool of polypeptides is
10 identified, DNA encoding each member polypeptide is incorporated into a corresponding
11 expression construct that produces the desired levels of protein expression. If it is desired to
12 adhere the polypeptides to a support (e.g., to a bead acting as either a location-determinable
13 support or as a randomizable support), then the DNA encoding each member polypeptide is
14 fused in frame with DNA encoding a suitable adhesion partner to form a
15 polypeptide/adhesion moiety fusion construct, described elsewhere herein. Optionally, as
16 described in more detail below, the construct may also utilize a downstream marker that
17 provides rapid indication of whether the fusion construct is in fact expressed in frame, and
18 with no premature terminations, and/or in a stable, suitably folded conformation.

19 In the case of screening the native cellular proteins of an organism, an expression
20 library is created by standard techniques, generating a sufficient number of fragments of
21 DNA so as to ensure that all protein domains are likely to be expressed in the library.
22 Sambrook, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor
23 Laboratory Press (1989), Chapters 7-9. Genomic DNA, cDNA synthetic or cloned DNA
24 sequences may be used. As one non-limiting example, synthesis of cDNA and cloning are
25 accomplished by preparing double-stranded DNA from random primed mRNA isolated from,
26 e.g., human placental tissue. Alternatively, randomly sheared genomic DNA fragments may
27 be utilized. In either case, the fragments are treated with enzymes to repair the ends and are
28 ligated into an expression vector suitable for expression in, e.g., *E. coli* cells. Exemplary
29 vectors include inducible systems, e.g., the trc promoter system, which is induced by addition
30 of suitable amounts of IPTG.

31 If a subcloning strategy is to be employed, the library polypeptide-encoding vectors
32 may be introduced into *E. coli* and clones are selected. Before proceeding with the inventive
33 method, the quality of the selected library optionally may be examined. For example, a set of

1 100 clones can be picked and sequenced at random, looking for homologies to known genes,
2 evidence of splicing, and such features. Alternatively, the library representation can be
3 explored by filter hybridization using probes of sequences of known abundance such as actin
4 and tubulin. These sequences should be present at a frequency in the library of between
5 0.01% and 1.0%.

6 Once a satisfactory polypeptide-encoding library (or, alternatively, DNA encoding a
7 desired set of individual polypeptides) is obtained, DNA encoding a suitable adhesion moiety
8 may be incorporated in frame with the polypeptide encoding DNA sequences. This DNA
9 fusion construct is then placed under control of a selected promoter in an expression vector
10 construct, so that upon induction one obtains suitably high levels of expression of the fusion
11 construct. There are many suitable adhesion moieties known to the art, including without
12 limitation biotin/avidin, thioredoxin/PAO, calmodulin binding peptide/calmodulin,
13 dihydrofolate reductase/methotrexate, maltose-binding protein/amylase, chitin-binding
14 domain/chitin, cellulose-binding domain/cellulose, glutathione-S-transferase/glutathione, or
15 antibody/antibody epitopes such as the FLAG epitope. One of ordinary skill may choose an
16 adhesion moiety that binds either reversibly or irreversibly to its complementary moiety. One
17 factor to consider in selecting an adhesion moiety complex is the relative spontaneous
18 dissociation constants (K_D) of the complexes. For example, the biotin/avidin link has a K_D of
19 approximately 10^{-15} M and is therefore relatively stable and irreversible. Maltose binding
20 protein/amylase, on the other hand, is less stable, with a K_D of 10^{-6} M. One option to increase
21 stability is to use cross-linking, for example by selecting a fusion protein with an adhesion
22 moiety that can be cross-linked by UV light.

23 The expression vector is chosen based largely on its ability to generate moderate to
24 high expression levels of either a given polypeptide or a fused polypeptide/adhesion moiety
25 (termed herein, a "fusion construct"), in a host cell of interest. *E. coli* is one such host cell,
26 although those of skill will appreciate that other bacterial, yeast or mammalian host cells, for
27 example 293 cells, are also suitable for use in the present invention. In the case of *E. coli*,
28 many suitable expression vectors are known to those in the art. For example, the expression
29 vector may employ the P_L , P_R , P_{lac} , P_{tac} , P_{trc} , P_{trx} or T7 promoters, to name only a few such
30 promoters known to those in the art. These promoters are regulated such that high level
31 expression is induced via increased growth temperature (from P_L or P_R through a mutant
32 temperature-sensitive form of the lambda repressor, cI857) or by addition of a suitable
33 inducing agent (e.g., IPTG for P_{lac} or P_{tac}) to the media. In order to provide a recognition

1 sequence for detecting interacting polypeptide/ligand pairs, the expression vector may
2 optionally be constructed to produce a fusion protein that consists of an N- or C-terminal
3 recognition domain (for example, an epitope that is specifically recognized by an antibody),
4 followed in frame by a sequence encoding the desired library polypeptide which is optionally
5 flanked by sites to facilitate cloning, followed by an N- or C-terminal adhesion domain to
6 enable attachment to a solid support, depending on the strategy employed.

7 Optionally, the expression vector may include a suitable downstream marker such as a
8 reporter or antibiotic resistance gene, by which one may determine whether the expression
9 vector construct is intact and correctly in frame. This variant includes in the above-described
10 DNA fusion construct an additional marker sequence designed to sort out viable constructs
11 from, e.g., out of frame or inverted constructs. Suitable reporter sequences include green
12 fluorescent protein, which is one of a family of naturally occurring fluorescent proteins
13 whose fluorescence is primarily in the green region of the spectrum, or modified or mutant
14 forms having altered spectral properties (e.g., Cormack, B.P., Valdivia R.H. and Falkow, S.,
15 *Gene* 173: 33-38 (1996)). (Both native GFP and such related molecules are collectively
16 referred to herein as "GFP") Alternatively, this GFP reporter may be inserted into the
17 expression construct in place of the adhesion domain if only the integrity of the library
18 polypeptide-encoding portion of the construct is of interest. Non-fluorescent markers of
19 construct integrity may also be employed, including a variety of antibiotic resistance genes
20 that are familiar to the art.

21 Fluorescent reporters such as GFP allow for subsequent rapid sorting of expression
22 products using flow cytometry with a fluorescence-activated cell sorter (FACS) machine.
23 This FACS sorting detects expression constructs that properly read through the GFP reporter
24 sequence and which are expressed at desirably high levels. Cells that express intact, in-frame
25 constructs are readily separated by detecting and collecting "bright" cells, which have an
26 intact GFP moiety that is properly in-frame with the polypeptide of interest, correctly folded,
27 and located downstream from a functional promoter. Constructs that are not intact will be
28 dim. Similarly, constructs with mutations or frame-shift deletions will eradicate the proper
29 relationship of the GFP moiety to the promoter, and the cells bearing such constructs will be
30 dim. Collecting only bright cells in this enrichment step significantly reduces the number of
31 underexpressed or nonfunctional fusion polypeptides that proceed into subsequent screening
32 steps. If antibiotic resistance is used as a marker, then transformed cells are plated on

- 1 antibiotic-bearing media; only those cells that read through completion of a construct that
- 2 includes an intact, downstream antibiotic resistance gene will survive and grow.

3 After the GFP-expressing clones are isolated, the polypeptide or fusion construct
4 inserts can be recovered. If the polypeptide library/adhesion moiety DNA fusion construct
5 was screened, the GFP reporter sequence may optionally be deleted from the vector using
6 standard restriction endonuclease fragment excision and religation, or other such techniques.
7 If only the library polypeptide-encoding constructs were screened but fusion to an adhesion
8 moiety is desired, then the polypeptide-encoding fragments are transferred into a vector
9 containing the adhesion-domain, or alternatively, the adhesion-domain-encoding sequence
10 can be inserted into the vector, or swapped into the vector in exchange for the GFP reporter
11 sequence. Other markers such as antibiotic resistance genes may similarly be removed, if
12 desired.

13 **Generation of individual arrays**

14 Next, each substrate must be individually arrayed at a unique location. In the case of
15 proteinaceous substrates, each corresponding clone is arrayed separately, in a unique location,
16 so that in subsequent steps, the identity of any particular polypeptide may be determined by
17 cross-referencing back to its unique location in the original array. Non-proteinaceous
18 substrates may be arrayed directly, without a preceding expression step.

19 In order to obtain a source of only a given single polypeptide, a single-cell clone is
20 obtained as follows. Once the above-described DNA fusion constructs are assembled,
21 selected host cells are transfected or transformed by standard gene transfer techniques such as
22 electroporation. The transformed cells are selected by growth of colonies on selective media
23 familiar to those of skill in the art (e.g., standard ampicillin-enriched Luria Broth). Single
24 colonies are then picked and placed into growth media in, e.g., 384-well microtiter trays. A
25 robot may be used for this purpose. If desired, duplicate trays may be prepared bearing host
26 cells of identical clones in identical array locations on a separate set of microtiter trays.
27 (Duplicate arrays are particularly desired if the ligands to be screened also are polypeptides --
28 i.e., if a protein-interaction map is sought).

1 As a result of this step, library arrays in, e.g., 384-well plate format are generated in
2 which each well produces a unique polypeptide (derived either from a library, or from a
3 selection of individual polypeptides of interest). Thus, in later steps, the identity of a
4 particular polypeptide may be determined by tracing its origin back to this corresponding
5 unique array location.

6 **Generation of lysate plates**

7 Once each desired polypeptide is being expressed by the corresponding host cells, the
8 cells are lysed so as to release the polypeptides. This growth and lysis may be accomplished
9 directly, in each unique array location that contains them (e.g., microtiter well).

10 Alternatively, in some embodiments each single-cell clone may be grown in an intermediate
11 location of larger size or volume, so that a greater number of cells may be generated and
12 concentrated for lysis. In such embodiments, each concentrated volume of polypeptide is
13 then either lysed and the lysate transferred to its corresponding, unique array location, or the
14 concentrate is transferred to that array location and then lysed *in situ*. Each clonal lysate then
15 is kept separate from every other, and in a unique location that can be referenced throughout
16 the ligand screening process. Thus, each soluble lysate can be correlated back to its unique
17 library array location, and the identity of the library polypeptide ascertained thereby, as the
18 soluble lysates are used in later ligand interaction screening steps.

19 In order to obtain the uniquely arrayed soluble lysates, the host cells first are grown
20 until mid- or late-log phase. Expression of the DNA fusion constructs (library polypeptide
21 and adhesion moiety) is induced by whatever method is required by the selected promoter
22 (e.g., IPTG or by raising the growth temperature to 42° C). After one to five hours of
23 continued cell growth under inducing conditions, the cells are lysed to free the library
24 polypeptide/adhesion moiety fusion constructs.

25 Any methods familiar to those of the art may be used to free the polypeptides of
26 interest from the host cells. For example, the host cells may be treated with lysozyme to
27 remove the cell wall, followed by hypotonic shock to disrupt the cell membranes and release
28 the contents of the cell into the buffer. The cells alternatively may be sonicated, lysed with a
29 freeze/thaw protocol, or lysed by addition of detergent. The lysate may optionally be
30 concentrated by standard techniques prior to further process steps. Alternatively, the library
31 polypeptide or its corresponding fusion construct may be secreted by the cells, in which case
32 the growth media rather than the cells are further processed.

1 **Other ligands**

2 In some embodiments of the invention, the screening may seek to identify interacting
3 pairs of endogenous polypeptides, in which case duplicate sets of soluble lysate arrays may
4 be generated from the same set of library polypeptides. In other embodiments, a variety of
5 other ligands may be tested for interaction with the original library polypeptides or other
6 substrates of interest. These other ligands may be proteinaceous in nature, in which case the
7 above procedure may be modified slightly so that a set of host cells expressing the
8 proteinaceous ligands is generated, and the corresponding array obtained.

9 In other cases, exogenous ligands may be screened for interaction with the
10 polypeptides of interest. Ligands such as small molecules, natural products, hormones,
11 receptors, antibodies, peptides, enzymes, growth factors, cellular adhesion molecules,
12 combinatorial library components and the like may be exposed directly to an appropriate
13 randomizable support (e.g., a support that will adsorb sufficient amounts of the ligand). In
14 other instances, the ligands may require initial derivitization so as to be chemically reactive
15 with surface functional groups on the support, in which case the ligands are, e.g., covalently
16 linked to the support. Alternatively, the ligands may be synthesized on the support.
17 Alternatively, this screening methodology can be altered slightly to serve as a displacement
18 assay, wherein a secondary ligand such as a small molecule is exposed to the primary
19 ligand/substrate pair. The secondary ligand may advantageously be adhered to a
20 randomizable support with a unique tag (for embodiments in which a large or very large
21 number of such secvondary ligands are screened). Alternatively, for embodiments in which a
22 lesser number of secondary ligands are screened, such secondary ligands can be free in
23 solution. In either event, pairs in which the secondary ligand displaces the primary ligand
24 can be detected, collected and analyzed as described elsewhere herein.

25 **Preparation of randomizable supports with unique tags**

26 In order to screen a variety of ligands for interaction with a given polypeptide, the
27 method generally requires using a support or substrate that will serve three functions; (a) it
28 will adhere to the ligand of interest; (b) it will be fully randomizable, so that an aliquot
29 containing a representative sampling of ligands may be presented to each polypeptide of
30 interest, and (c) it will carry a unique identification tag that corresponds to the particular
31 ligand adhered to its surface, and distinguishes it from other ligand-bearing supports.

32 In one embodiment of the invention, the randomizable support is a bead or other such
33 microparticle. A variety of bead sizes and compositions are suitable for use in the present

1 invention. For example, bead size may range from 50 nm to 50 microns in diameter. The
2 beads may be composed of polystyrene, glass (silica), latex, agarose, magnetic resin, or a
3 variety of other matrices. Some beads may be obtained from commercial sources with
4 adhesion moieties already attached; for example, numerous avidin-conjugated beads are
5 available. Other beads can be obtained with functional groups such as hydroxyl or amino
6 groups suitable for chemical modifications, such as attachment of adhesion moieties that will
7 interact with the fusion protein. In yet another formulation, the beads do not require specific
8 functional groups; rather, the interaction between the fusion protein and the bead is of a
9 nonspecific type involving, e.g., hydrophobic interactions. Beads suitable for this purpose
10 may be polystyrene, latex, or some other plastic.

11 If the beads require functionalization in order to bind to the selected polypeptide or
12 ligand, then enough beads are generated in one reaction to permit numerous experiments to
13 be performed, e.g., 10^{14} beads. These beads are then stored under conditions that ensure the
14 stability of the chemical modifications, such as low temperature. For example, in mapping
15 protein interactions in a human cell, approximately 1×10^7 beads are generated for each
16 potential expression product to be screened (e.g., in the case of the human cell, approximately
17 1×10^6 potential endogenous polypeptides, resulting in a need for some 1×10^{13} beads). This
18 number of beads ensures that at least one full experiment involving genome-wide protein-
19 protein interaction measurements can be performed.

20 A variety of methods are suitable for providing each support with an identification tag
21 that correlates to the ligand that the support will bear. For example, the beads may be tagged
22 with DNA tags in which the tags can be amplified and fingerprinted, or detected by
23 hybridization. Alternatively or in conjunction, the beads may be tagged with fluorescent tags
24 such as fluorescent barcodes, radio frequency tags, or mass tags detected by mass
25 spectrometry

26 **Fluorescent barcodes**

27 Fluorescent tags for the randomizable supports are advantageous because the
28 identification tag may be read simultaneously with quantification of the binding interaction.
29 One representative method of fluorescent tagging is to use the variety of existing fluorescent
30 materials such as fluorescent organic dyes or microparticle dyes, and the sensitivity of
31 existing fluorescence detectors, to devise a series of fluorescent barcodes.

32 Fluorescent barcodes may be generated as follows. Fluorescence detectors presently
33 exist that can quantify fluorescence at up to nine separate wavelengths using multiple lasers,

1 photo-multiplier tubes (PMTs) and filter sets. One example of such a device is the
2 Cytomation flow cytometer that is not only capable of measuring fluorescence at multiple
3 wavelengths in single cells or beads, but also of sorting cells and beads based on these
4 signals. The measurements are also highly accurate, so that it is possible to distinguish easily
5 a fluorescence value of 0 (background) from, 1x, 2x, 3x, and 4x. Thus, it is possible to
6 design a barcoding strategy whereby the unique signature of a particular bead is based on a
7 fluorescence number composed of, e.g., nine digits (i.e., the nine separate wavelengths), each
8 digit able to assume 5 values (i.e., 0 through 4x). Combining these two variables yields a set
9 of potential unique barcodes of 5^9 , or approximately 2 million different barcodes.

10 To stamp each bead with a barcode, a set of, e.g., 1×10^{13} beads is broken into one
11 million groups of 1×10^7 each. Each group of beads is placed in one well of a 384-well tray,
12 requiring a total of about 2,600 trays. As one of skill will appreciate, this process may
13 preferably be automated via known methods, using commercially available robotics. To the
14 beads are added various quantities and types of fluorochrome dye such that the barcode
15 requirements are fulfilled -- i.e., that each type of bead has a unique barcode that will identify
16 the associated ligand and distinguish it from all other ligands. The fluorochromes may
17 readily be incorporated by dissolution in organic solvent followed by exposure to the beads
18 for sufficient time to allow full diffusion and interaction with the beads. The organic solvent
19 is then removed and the beads dried. Alternatively, various types of covalent chemical
20 attachments to the beads may be employed, or the fluorescent dye may be incorporated into
21 the bead by other methods known to the art, for example by synthesizing the beads from dye
22 containing materials, or by encapsulating the fluorescent dye within the bead.

23 Generation of a randomized ligand library for screening

24 Once the beads are prepared with the desired fluorescent barcode or other such
25 unique tag, the desired ligands (or secondary ligands) may be adhered to the beads, to form a
26 series of uniquely tagged ligand sets.

27 A variety of methods for adhering a ligand to the support are known to the art, and
28 one of ordinary skill can select a particular method based on the exact nature of the ligand to
29 be adhered. For example, if the ligand is proteinaceous, the adhesion moiety may be, e.g.,
30 biotin/avidin, thioredoxin/phenyl arsine oxide, maltose binding protein/amylose,
31 calmodulin/calmodulin binding peptide, dihydrofolate reductase/methotrexate, chitin/chitin
32 binding protein, cellulose/cellulose binding protein or antibody/antibody epitopes such as the
33 FLAG epitope, as described elsewhere herein. In each case, one binding moiety is expressed

1 as part of a fusion construct in frame with the proteinaceous ligand, and the other is
2 immobilized on the support by a covalent or noncovalent chemical linkage. In the case of
3 hormones or other endogenous compounds, or other organic or inorganic molecules, the
4 compounds may be attached via a chemical linker, e.g., a hydroxyl or primary amine, or may
5 be synthesized directly on the bead.

6 If the ligand to be adhered is proteinaceous, then a subset of uniquely tagged,
7 derivatized beads is exposed to a corresponding expression product lysate, which is collected
8 in a particular location in, e.g., a 384 well array. The subset of identically tagged beads is
9 suspended in solution and added to each well by either a pipetting device or by means of a
10 magnetic dispenser (in the event that the beads are magnetic). The beads are mixed with the
11 lysate in the well for a sufficient time to permit binding. This step thus generates subsets of
12 uniquely identified ligands on randomizable supports.

13 It is most preferable to adhere each member ligand to its corresponding set of
14 location-determinable supports in a substantially irreversible manner. Some adhesion
15 moieties form such links by a covalent link or an extremely tight noncovalent link -- e.g., the
16 interaction between biotin and avidin, $K_d = 10^{-15}$ M. Such substantially irreversibly linked
17 beads are ready for the next step in the process -- exposure of the substrates to ligands that are
18 firmly bound to their randomizable supports. However, if the interaction between the
19 randomizable support and the ligand is reversible (e.g., on the order of $K_d = 10^{-6}$ to 10^{-10} M),
20 an additional step may be employed. In this additional step, the ligands are eluted from the
21 first set of supports (which may, in this instance, be unlabelled, as the various subsets of
22 ligands at this juncture remain segregated) by addition of a large excess of soluble (i.e.,
23 unbound) ligand.

24 In the case of polypeptide/adhesion moiety fusion constructs, one adds an excess
25 soluble adhesion moiety so as to competitively interfere with the interaction between the bead
26 and the adhesion domain of the fusion construct, thus displacing the fusion construct from the
27 bead. The soluble fusion construct then is re-attached via an irreversible linkage to another
28 set of beads that are added to the solution in a location-determinable manner. This interaction
29 may involve, e.g., binding avidin-coated beads by biotinylated fusion protein, or it may
30 involve nonspecific, hydrophobic adsorption of the soluble protein onto the bead surface.
31 Alternatively, it may be preferable to crosslink polypeptides to beads using, e.g., UV light of
32 a specific wavelength and/or a chemical cross-linking agent, as is the case with the
33 randomizable supports, described elsewhere herein.

Once all subsets of uniquely tagged beads have been successfully linked to the corresponding ligand subsets, then all the ligand subsets are collected by either a pipetting device or by the magnetic instrument and mixed into one integrated pool such that, e.g., all 1 $\times 10^{13}$ ligand-labeled beads are present. This step thus disperses all the tagged ligands into a fully randomized pool that represents all of, e.g., the one million protein-bead types, each type represented 10⁷ times. Each bead in the aliquot bears a ligand and a corresponding unique tag to identify that ligand. An aliquot of, e.g., 10⁷ beads is then drawn from this integrated pool of ligand-bearing beads. Each aliquot contains a statistically representative portion of the fully integrated ligand pool -- i.e., a subset of beads representing a substantially full spectrum of available ligands (the degree of complete representation in any selected aliquot is determined by statistical sampling issues familiar to those in the art). Each location in the substrate array receives one aliquot of integrated ligand beads. Thus each arrayed substrate has the opportunity to interact with every ligand.

Preparation of a location-determinable support and exposure to substrates

Alternatively, in some embodiments of the invention, the substrates are adhered to a location-determinable support prior to exposure to the aliquots of integrated ligand-bearing supports. Generally, the two major characteristics of the location-determinable support are that (i) it is capable of adhering to the selected library polypeptide or other such substrate, and (ii) it is kept segregated so that it links the adhered substrate to the original clone array position (i.e., well) from which that substrate was derived. This support can be a fixed type of support, for example a finger, pin or other such probe that is rigidly arrayed so as to match the clone array (e.g., a 384 pin hand). Alternatively, the support can be a bead or other such microparticle, which is kept segregated in an array that directly correlates back to the original location in the substrate array (e.g., a set of beads that is kept segregated in one well of a 384 well tray, corresponding to the well of the 384 well tray from which, e.g., the original clonal polypeptide was derived). Microparticles may be preferable for selections that involve large numbers of substrate-ligand interactions, or that involve relatively specific or slow-forming interactions. Fixed supports offer advantages for reduced handling and/or automation.

As described above, it is most preferable that the substrate be linked in a substantially irreversible manner to the location-determinable support. If this is not accomplished by the initial adhesion step, then the substrates are eluted from the first set of supports by addition of a large excess of soluble (i.e., unbound) substrate. The substrate is then re-adhered to a

1 second set of location-determinable supports in a substantially irreversible manner, as
2 described above.

3 **Exposure of each substrate to the integrated ligand library**

4 Generally, this step requires that each uniquely located substrate (either in solution or
5 adhered to its analogous location-determinable support) is exposed to an aliquot of integrated
6 ligand-bearing supports. Typically, these ligands will be in an appropriate buffer that mimics
7 conditions inside the cell (i.e., reducing environment, neutral pH, 150 mM salt), and can be
8 added directly to each array location containing a corresponding soluble or bound substrate.

9 The lysate buffer may be of the same makeup. The binding buffer also may have other
10 additives, e.g., those designed to minimize non-specific binding (e.g., detergent, bovine
11 serum albumin). If a fixed type of location-determinable support (e.g. a pin or finger) is used,
12 it may simply be dipped into a well containing an aliquot of the randomized ligand-bearing
13 supports. If the location-determinable support is a bead or other such microparticle, a set of
14 such beads containing one particular substrate may be added to a well that contains a
15 randomized aliquot of the ligand-bearing beads, and the two sets of beads mixed thoroughly
16 so as to maximize substrate-ligand exposure. Interaction between the substrate and any of the
17 many different ligands thus results in the corresponding ligand-bearing bead (with its unique
18 identification tag) adhering to the substrate, thereby forming a bead-bead aggregate.

19 In some embodiments utilizing microparticles as location-determinable supports, it
20 may be desirable to replace the support-bound substrate with soluble substrate after exposure
21 to the ligand aliquots (and formation of substrate-ligand bead aggregates). In such cases,
22 soluble substrates (termed herein, "replacement substrates") are added to each array location
23 that contains the corresponding bead aggregates. For example, in the case of individual or
24 library polypeptides, the polypeptide domains of the replacement polypeptides are identical to
25 those of the polypeptides bound to the supports. Because the replacement polypeptides are in
26 vast excess, and because the interactions between polypeptides and ligands in solution are
27 generally characterized by relatively rapid off-rates, the soluble replacement polypeptides
28 bind the ligands and displace competitively the support-bound polypeptides. Thus, in a
29 single step the location-determinable supports are displaced from the ligand-bearing
30 randomizable supports and soluble replacement polypeptides are attached to the ligand-
31 bearing supports in preparation for further characterization or screening. For example, in
32 embodiments in which both the replacement substrate and the ligand are proteinaceous, the
33 pairs may be subsequently exposed to secondary ligands, typically small organic molecules,

1 as described herein. Small organic molecules that bind to the primary ligand, for example,
2 can displace the replacement substrate, thereby identifying small a organic molecule with
3 potential therapeutic value as a disruptor of a protein-protein interaction.

4 Alternatively, it may be preferable to detach the location-determinable supports in a
5 separate step, followed by incubation of the segregated sets of interacting ligand-bearing
6 beads with soluble replacement polypeptide or such substrate. This may be accomplished,
7 for example, by hydrolysis of a linker that attaches the library polypeptides to the location-
8 determinable supports. If a DNA linker is used, DNase treatment may release the location-
9 determinable beads, while the residual fusion protein remains bound by noncovalent forces to
10 the ligands on the randomizable beads. A second binding step involving the ligand-bearing
11 beads and soluble replacement polypeptides is then performed in order to adhere the second
12 layer (the library polypeptide layer) to the bead prior to detection of polypeptide-ligand
13 complexes. This replacement step is generally applicable to non-proteinaceous substrates, as
14 well.

15 **Magnetic interactions**

16 In one embodiment of the invention, beads formed from a magnetic resin are used as
17 the location-determinable support. In this embodiment, a set of magnetic beads (e.g., 10^7
18 beads per well) is apportioned into each array location, which contains a corresponding
19 library polypeptide or other such substrate. As the magnetic beads have adhesion domain
20 binding moieties that are complementary to those of, e.g., the fusion polypeptides conjugated
21 to their surfaces, after some period of time saturating or near-saturating amounts of fusion
22 protein will adhere to the resin, and the polypeptide-coated beads are collected. This may be
23 accomplished by dipping a magnetic pin into each well, allowing the magnetic beads (with
24 the adhered substrates) to be drawn to the pin, withdrawing the beads, transferring to another
25 well, and discharging the magnetic bead by demagnetizing the pin. In other embodiments,
26 the magnetic forces may be applied externally to pull the magnetic beads to the well wall,
27 with subsequent removal of the remaining non-magnetic materials.

28 Next, substrate/ligand bead aggregates are formed and collected. First, each set of
29 magnetic beads in the array is exposed to aliquots of non-magnetic ligand-bearing supports.
30 After a period of time to permit interactions between substrates and ligands, the magnetized
31 beads are again collected with the aid of a magnetic device. Any of the ligand-bearing beads
32 that have interacted to form aggregates with the magnetized beads are pulled along with the
33 magnetic beads to the magnet. Ligand-bearing beads that do not interact are left behind in

1 solution. The aggregates of magnetic beads and interacting ligand-bearing beads are then
2 collected. Thus, only those beads that contain interacting substrates and ligands are
3 recovered for subsequent quantitative analysis.

4 Conversely, the ligand-bearing randomizable supports may be magnetized while the
5 location-determinable supports remain unmagnetized. The magnetized randomizable
6 supports then function analogously to gather the bead aggregates formed by the
7 substrate/ligand complexes.

8 In using magnetic forces to cull out interacting substrate/ligand complexes, a "surface
9 interaction" as opposed to solution interaction is created, and provides an enrichment for
10 substrate-ligand interactions. This enrichment step obviates the need to examine carefully
11 every possible substrate-ligand interaction using a quantitative, but serial device such as a
12 flow cytometer. Accordingly, interaction sets on the order of $10^6 \times 10^6$ polypeptides (akin to
13 a human protein interaction map) may be screened rapidly and efficiently by inserting a bead-
14 bead interaction step.

15 **Segregating, identifying and quantifying the substrate/ligand pairs**

16 Once the substrate/ligand interactions are consummated, the interactions can be
17 quantified, and each substrate and ligand identified as follows.

18 In the case of proteinaceous substrates, one ultimately obtains a set of supports that
19 bear a polypeptide layer reversibly bound to ligand-bearing randomizable supports (i.e.,
20 either the randomizable supports were exposed only to soluble polypeptides, or the bead-
21 bound polypeptides were subsequently displaced by an intervening exposure to soluble
22 polypeptides). Such polypeptide/ligand complexes may be rapidly quantified by use of a
23 fluorescence-activated cell sorter. The fluorescent signals emitted by the unique tags on the
24 ligand-bearing supports provide the basis for rapid and accurate quantitation by this method.

25 In other embodiments, substrate-ligand complexes can be detected by either detecting
26 a unique recognition domain (e.g., epitope) on the polypeptide or ligand (by "unique" is
27 meant either that the recognition domain exists on only one member of the complex, or
28 alternatively that it is present on both members but sterically accessible only on the outer
29 layer). Supports that bear a ligand may be identified by a variety of immunological or
30 fluorescence techniques known to those in the art. As one non-limiting example of such
31 identification, a fluorescence-labeled antibody that reacts with such an epitope on the library
32 polypeptide is utilized. After a period of time suitable for antibody binding (typically one
33 half hour), the beads are collected and examined by an instrument such as a FACS machine

1 to measure the level of antibody (determined from the fluorescence signal of the particular
2 fluorochrome attached to the antibody). Concurrently, the randomizable support barcode can
3 be read by fluorescence measurements at other wavelengths. This in turn reveals the identity
4 of the fusion protein attached irreversibly to the randomizable support. The identity of the
5 soluble protein is retained based on the well from which the bead was collected (i.e. the
6 unique array location) immediately prior to the detection step. Thus, both the identity of the
7 primary, irreversibly attached protein and the soluble protein is known, and the approximate
8 strength of the interaction between them can be determined from the antibody fluorescence
9 signal.

10 For some applications, a CCD camera may be utilized to detect interacting substrate-
11 ligand complexes. For example, in applications screening for interaction of a non-
12 proteinaceous organic molecule with a polypeptide, a CCD system can be used to visualize
13 interacting complexes, thereby providing both detection and quantification. The CCD
14 camera can detect a variety of visual outputs, including without limitation fluorescent
15 emissions, chemiluminescent emissions, and SPA (scintillation Proximity Assay) emissions.
16 In the SPA format, one member of the interacting pair is radiolabeled using standard
17 techniques, and the other member of the pair is adhered to a bead in which a radio-detecting
18 scintillation component is incorporated in the interior of the bead. When the radiolabeled
19 component interacts with the bead-bound component, a detectable scintillation signal is
20 emitted. The beads can optionally be displayed on some surface, for example an
21 identification grid with grid locations correlating to each unique array location, for scanning
22 by the detector.

23 One non-limiting example of CCD detection of fluorescent signals utilizes a scientific
24 grade CCD camera incorporating a high quantum efficiency image sensor. The target
25 molecules are distributed along the well bottoms of optically transparent microtiter plates.
26 The CCD, fitted with lenses and optical filters, acquires images of the through the optically
27 transparent well bottoms. Fluorescent excitation of the fluorescent molecules is generated by
28 appropriately filtered coherent or incoherent light sources. The resulting digital images are
29 stored on a computer for subsequent analysis.

30 An exemplary detection system is composed of a PixelVision SpectraVideo™ Series
31 imaging camera (1100 x 330 back-illuminated array), PixelVision PixelView™ 3.03
32 software, two 50-mm/f1.0 Canon lenses, four 20750 Fostec light sources, four 8589 Fostec
33 light lines, one 59345 Oriel 510-nm band pass filter, four 52650 Oriel 488-nm laser band pass

1 filters, a 4457 Daedal stage, Polyfiltronic clear bottom microtiter plates, and supporting
2 mechanical fixtures. Mechanical fixtures are constructed to position the PixelVision camera
3 below a microtiter dish. Additionally, the fixtures mounted four Fostec light lines and
4 allowed the excitation light to be focused on the viewed area of the microtiter dish. The two
5 Canon lenses were butted up against each other front to front. A 510-nm filter is placed
6 between the two lenses. The front-to-front lens configuration provides 1:1 magnification and
7 close placement of the target object to the imaging system.

8 The above-described techniques quantify polypeptide binding pairs or
9 polypeptide/ligand binding pairs. Optionally, the exact make-up of each binding pair is
10 ascertained by identifying (i) the unique array location from which the library polypeptide or
11 other such substrate is derived, and (ii) the ligand identity that corresponds to the unique tag
12 on the bead (which, in the case of creating protein interaction maps, will in turn relate back to
13 another unique library polypeptide array location). Optionally, if sequence information about
14 a given interacting polypeptide is desired, one may sequence the DNA encoding the
15 polypeptide produced by each unique location in the library array.

16

DESCRIPTION OF PREFERRED EMBODIMENTS

18

EXAMPLE 1

20

LYSATE LIBRARIES

Expression vectors

22 In order to generate sufficient amounts of polypeptides for ligand screening, it is
23 desirable to first clone DNA encoding the library polypeptides of interest into a vector that is
24 suitable for high levels of expression of those polypeptides. The host cells of interest are
25 transformed with such an expression vector, production of the library polypeptides is
26 induced, and the library polypeptides are collected.

27 A variety of expression vectors are suitable for use in this invention. As one non-
28 limiting example, an expression vector bearing an inducible trc promoter was used. Plasmid
29 pSE420 (Invitrogen) features the trc promoter, the lacO operator and lacI^q repressor, a
30 translation enhancer and ribosome binding site, and a multiple cloning site. For insertion into
31 this vector, the *E. coli* thioredoxin gene was amplified from pTrx-2 (ATCC) in such a manner
32 as to retain a restriction enzyme site on the 5' side of the gene, and was cloned into the
33 pSE420 vector's multiple cloning site at the 5' NheI and 3'NgoMIV locations, thus placing it

1 under control of the trc promoter. The thioredoxin gene can advantageously enhance
2 recombinant protein solubility and stability. Moreover, as a cytoplasmic protein, it can be
3 produced under reducing conditions but still can be released by osmotic shock because of
4 accumulation at adhesion zones.

5 Once the pSE420 plasmid was modified to contain the thioredoxin gene
6 (pSE420/trxA), the gene encoding GFP was inserted in frame with the thioredoxin, in order
7 to rapidly isolate intact, in-frame constructs and thereby to eliminate constructs in which the
8 library polypeptide would not be properly produced. The gene encoding EGFP was PCR
9 amplified from plasmid pEGFP-1 (Clontech), maintaining a NotI restriction site 3' of the
10 EGFP sequence, and establishing a second NotI site 5' of that sequence. The NotI sites may
11 be used to readily remove the EGFP fragment from the vector after intact constructs are
12 isolated. The NotI fragment containing EGFP was then cloned into the NotI site of the
13 pSE420/trxA vector. Vectors containing the EGFP in frame and in the correct orientation
14 were designated plasmid pSE420/trxA/EGFP. Figure 1.

15 Once the vector containing the desired promoter and other components is prepared,
16 DNA encoding the desired adhesion moiety is introduced. For example, a biotinylation
17 signal may be used to adhere the library polypeptides to streptavidin beads. The *in vivo*
18 biotinylation peptide sequence was cloned into the pSE420/trxA/EGFP vector (Figure 1) in
19 frame to the amino terminus of the thioredoxin gene by cutting at the 5' NcoI and 3' NheI site
20 and filling in the overhanging nucleotides with Klenow prior to ligation. The biotinylation
21 signal peptide is 23 residues long (Tsao et al, *Gene* 169:59-64 (1996)), and the sequence that
22 encodes it can be readily synthesized on an oligonucleotide synthesizer using standard
23 techniques. The vector may advantageously be modified to include the BirA gene, which
24 encodes the enzyme responsible for adding biotin to the recombinant biotinylation signal.
25 The BirA gene was amplified from genomic *E. coli* DNA by PCR. A copy of the BirA gene
26 was added in a polycistronic fashion to the carboxyl terminus of the biotin/trxA/EGFP
27 sequence and the resultant modified pSE420 vector was designated pSE420/biotrx/GFP/BirA
28 (Figure 2).

29 An alternative adhesion moiety, dihydrofolate reductase (DHFR) was incorporated
30 into the expression construct as follows. The DHFR gene was amplified from *E. coli*
31 genomic DNA by PCR with NcoI and KpnI sites on the 5' and 3' ends, respectively. This
32 fragment was cloned into the NcoI/KpnI site of pSE420. Subsequently, the NotI fragment

1 containing EGFP (described above) was cloned in frame with DHFR into the NotI site. The
2 resultant plasmid was designated pSE420/DHFR/GFP (Figure 4).

3 Another promoter system suitable for use in the invention features the P_L promoter.
4 This system was constructed by digesting the pLex plasmid (Invitrogen) with NdeI and PstI
5 and blunting the resultant ends with mung bean nuclease. The pSE420/biotrxGFP/BirA
6 construct described above was digested with NcoI and HindIII, and the NcoI/HindIII
7 fragment then blunt-ended with T4 polymerase. This fragment was then inserted into the
8 pLex construct. The resulting plasmid was designated pLex/biotrx/GFP/BirA (Figure 5).
9 Optionally, the DHFR/GFP expression cassette described above may be inserted into the
10 pLex plasmid by digesting pLex with NdeI and PstI, blunting the ends with mung bean
11 nuclease, and inserting the blunt-ended NcoI/HindIII fragment from pSE420,DHFR/GFP.

12 Following construction of the described vectors, expression was induced by
13 introduction of the appropriate induction agent (IPTG for pSE420-based expression vectors,
14 and tryptophan for pLex-based vectors). Production of the recombinant polypeptide insert
15 was detected by GFP fluorescence via FACS, or by western blot analysis. The recombinant
16 polypeptides were then selectively bound and removed from bacterial lysates of induced
17 cultures via binding with the respective binding partner (streptavidin for biotrx/GFP and
18 methotrexate for DHFR/GFP), which had been immobilized to beads, as described elsewhere
19 herein.

20 *Library polypeptides*

21 DNA encoding the library polypeptides may be derived from a variety of sources, using
22 techniques that are familiar to the art. As one non-limiting example, a cDNA library
23 encoding human protein domains was prepared, using methods that are well known in the art,
24 from human placental tissue. Poly(A) RNA was isolated from placental tissue by standard
25 methods. First strand cDNA was then generated from poly(A) mRNA using a primer
26 containing a random 9mer, a SfiI restriction endonuclease site and a site for PCR
27 amplification (5'-

28 ACTCTGGACTAGGCAGGTTCAGTGGCCATTATGCCNNNNNNNN). The second
29 strand was then generated using a primer consisting of a random 6mer, another SfiI site, and a
30 site for PCR amplification (5'-
31 AAGCAGTGGGTCAACGCAGTGAGGCCGAGGCAGGCCNNNNNN). After conducting
32 a number of PCR amplification cycles, the DNA was cut with SfiI and the resultant fragments
33 were size-selected for fragments of greater than about 400 bp. The selected fragments were

1 ligated into the SfiI sites of a suitable expression vector, as described herein. The library
2 polypeptide DNA fragments then were isolated and inserted in frame with DNA encoding a
3 corresponding biotin adhesion moiety and thioredoxin. DNA encoding the library
4 polypeptides was prepared by cutting the DNA with SfiI and then inserted at an SfiI site
5 placed in a linker (5' GCCGAGGCGGCCTGATTAAACGATGGCCATAATGGCC) placed
6 at the NgoMIV-AvrII sites of plasmid vector pSE420/biotrx/GFP/BirA, or of plasmid vector
7 pET-biotrx-GFP-BirA.

8 To select for those cDNAs that are in-frame with TrxA, *E. coli* expressing constructs
9 possessing in-frame cDNAs are selected by FACS sorting and selecting for bright (i.e.,
10 "green") cells. Such cells are expressing intact GFP, which is in frame with and downstream
11 from the library polypeptide and TrxA sequences. Plasmid DNA is isolated and the EGFP
12 insert then removed via NotI digestion. Once the EGFP marker has been used to sort cells
13 and removed from the modified pSE420 vector, the modified pSE420 plasmids are again
14 transformed into *E. coli* and expressed via IPTG induction.

15 *Other adhesion moieties*

16 Alternatively, the library polypeptides may adhere to calmodulin-containing beads
17 using calmodulin binding peptide ("CBP") as the adhesion moiety. The vector constructs are
18 prepared as described above, but an expression cassette containing CBP is inserted into the
19 vector immediately 5' of the trxA gene via the 5' NcoI and 3' NheI sites, as described above.
20 Figure 3. The CBP thus is used in place of the biotinylation signal peptide, and immobilizes
21 the library polypeptides to the calmodulin beads.

22 As another alternative to the above-described system, the thioredoxin gene product
23 may itself serve as the adhesion moiety, and will bind the fused library polypeptides to
24 phenylarsine oxide ("PAO") beads. Polystyrene beads are modified so as to covalently link
25 phenylarsine oxide to the surface by reacting the carboxyl groups on the bead surface with p-
26 aminophenylarsine oxide via a water soluble carbodiimide. Kaleef and Gitler, *Methods of*
27 *Enzymology* 233:395-403 (1994). The above-described pSE420/trxA/EGFP vector in this
28 instance is used directly, i.e., no subsequent moiety is fused to the carboxyl terminus of the
29 thioredoxin gene. Screening and expression are carried out as described above.

30 As still another alternative, the library polypeptides may simply be adhered to
31 polystyrene beads via hydrophobic adsorption. In such embodiments, the library
32 polypeptides are first separated from, e.g., the host cell polypeptides by standard methods
33 before exposure to the beads.

1 *Crosslinked embodiments*

2 In some embodiments, polypeptide substrates or ligands may be crosslinked with the
3 supports. As one non-limiting example, the bacterial lysate containing the expressed
4 recombinant fusion protein is incubated with microspheres containing a ligand specific for
5 the fusion partner. Following binding of the fusion protein, a photoactive crosslinker on the
6 microsphere will irreversibly bind the fusion protein. Examples of possible ligand-fusion
7 partner combinations are, but not limited to, phenylarsine oxide (PAO) and thioredoxin
8 (Methods of Enzymology (1994) 233, 395-403), or a suicide substrate and its corresponding
9 enzyme (e.g. clavulanic acid and beta-lactamase; J. Mol. Biol. (1994) 237, 415-422).

10 In embodiments utilizing PAO and thioredoxin, the thioredoxin fusion product is
11 constructed as described above. The PAO moiety, 4-aminophenylarsine oxide, is synthesized
12 as described in the literature (Biochemistry (1978) 17, 2189-2192). The 4-aminophenylarsine
13 oxide is then reacted with a large molar excess of BS₃ (Pierce Chemical Co.) in order to place
14 an amine reactive NHS ester and 8 carbon spacer at the 4 position of 4-aminophenylarsine
15 oxide. The NHS ester-modified PAO is then reacted in equimolar amounts with sulfo-
16 SANPAH (Pierce Chemical Company) and 10 µm amine-functionalized latex microspheres
17 (Polysciences, Inc.). The result of this reaction yields microspheres with approximately one-
18 half of the available amine groups with PAO attached, while the remaining half have the
19 photoactivatable crosslinker. These microspheres are then reacted with the bacterial lysate
20 containing the expressed fusion protein. Vicinal dithiol-containing proteins, including the
21 recombinant thioredoxin fusion protein, is bound to the microspheres. After washing steps to
22 remove non-specifically bound proteins, the microspheres with the bound recombinant fusion
23 protein are crosslinked to the microspheres via amine groups on thioredoxin by exposing to
24 light at 320nm-350nm. These microspheres are then ready to be used as described elsewhere
25 in this application.

26 In another non-limiting embodiment, library polypeptides are covalently attached to
27 the supports by adsorption to the support, followed by crosslinking. For example, the library
28 polypeptides may be constituted as fusions with maltose binding protein. These fusion
29 constructs then are purified from the lysate using a maltose affinity resin and released with
30 soluble maltose (J. Chrom. 633 (1993) p.273-280). The purified fusion constructs then are
31 adsorbed onto polystyrene beads, thus attaching via hydrophobic interactions. Finally, the
32 polypeptides are crosslinked with a phototactivated crosslinker, for example sulfo-SANPAH
33 (Pierce Chemical Co.).

1 In yet another non-limiting embodiment, polypeptide substrates are attached to
2 microparticles via the interaction of a DNA-binding protein and a DNA moiety or analog on
3 a bead. Specifically, a DNA binding fusion library such as a Gal4 fusion is constructed. The
4 corresponding microparticles have two features -- a peptide nucleic acid (PNA) oligomer for
5 binding the protein of interest, and a photoactivatable crosslinker, e.g. sulfo-SANPAH
6 (Pierce Chemical Company), attached to the end of the oligomer. The microparticles are
7 placed into lysates containing the various Gal4/library polypeptide fusion constructs, and
8 those constructs then bind to the beads via interaction between the Gal4 binding moiety and
9 the bead oligomer. The crosslinker is then photoactivated, thus forming the covalent linkage
10 between the proteins and the beads.

11 Alternatively, the bacterial lysate containing the expressed recombinant fusion
12 polypeptides are incubated with microspheres that bear a ligand specific for the fusion
13 polypeptide. After the polypeptides bind to the beads via the ligands, a photoreactive
14 crosslinker on the bead is activated so as to irreversibly bind the fusion polypeptide to the
15 bead. Non-limiting examples of fusion polypeptide/ligand partners include
16 DHFR/methotrexate, PAO/thioredoxin, or a suicide substrate and corresponding enzyme
17 (e.g., clavulanic acid and beta-lactamase; *J. Mol. Biol.* (1994) 237:415-422).

18 For an embodiment utilizing the thioredoxin construct described elsewhere herein, 4-
19 aminophenylarsine oxide is synthesized as described in the literature (*Biochemistry* (1978)
20 17:2189-2192), reacting the 4-aminophenylarsine oxide with a large molar excess of BS³
21 (Pierce Chem. Co.) in order to place an amine reactive NHS ester and eight carbon spacer
22 at the 4 position of the 4-aminophenylarsine oxide. The NHS-modified PAO is then reacted
23 in equimolar amounts with sulfo-SANPAH (Pierce Chem. Co.) and 10 µm amine-
24 functionalized latex microspheres (Polysciences, Inc.), yielding microspheres with
25 approximately one half of the available amine groups with PAO attached, while the
26 remaining half attaches the photoactivatable crosslinker. The microspheres are then reacted
27 with the bacterial lysate containing the expressed thioredoxin fusion protein. Vicinal dithiol
28 containing polypeptides, including the recombinant thioredoxin fusion protein, are thus
29 bound to the microspheres. After washing steps to remove the non-specifically bound
30 protein, the microspheres with the bound recombinant fusion polypeptide are crosslinked via
31 the thioredoxin amine groups by exposing the complexes to 320-350 nm light.

32 For a DHFR/methotrexate embodiment, the DHFR expression vector is as described
33 elsewhere herein. The corresponding affinity resin, sulfo-SANPAH (Pierce Chem. Co.) is

1 reacted with the amine-functionalized latex microspheres (Polysciences Inc.) in non-
2 saturating amounts to couple the crosslinker onto the microspheres in non-saturating
3 amounts. Methotrexate (Sigma Chem. Co.) is then reacted with EDC (Pierce Chem. Co.) and
4 the sulfo-SANPAH functionalized beads so as to couple the methotrexate to available amine
5 groups on the beads. The resultant functionalized microspheres are depicted in Figure 6. A
6 bacterial lysate containing DHFR fusion polypeptide is then bound and photo-crosslinked as
7 described for the thioredoxin/PAO system.

8 In embodiments that utilize fluorescent identification tags, it may be preferable to first
9 protect the fluorescent tags before undertaking chemical cross-linking. This may be
10 accomplished in a variety of ways familiar to the art, including without limitation embedding
11 the fluorescent tags beneath the surface of the bead, or chemically protecting the fluorescent
12 tags by first derivatizing with non-reactive functional groups, and then de-protecting the tags
13 once chemical crosslinking is complete.

14 *Host cells*

15 A variety of host cells are suitable for use in this invention. One common species of
16 host cell with utility here is *E. coli*. Preferred strains of *E. coli* are characterized by (1) over-
17 expressing the necessary amount of protein required to fulfill other parts of the invention
18 (coating of the beads, etc.), (2) tolerating "leaky" expression of toxic target plasmids, and (3)
19 being amenable to cell lysis and protein recovery. Such strains include, without limitation,
20 TOP10 (Invitrogen Corporation), BL21 (Novagen), and AD494 (Novagen). One such strain,
21 BL21 (DE3) R/L (Stratagene), was selected for further study in this non-limiting Example.

22 These host cell strains are used in the presence or absence of the T7 phage gene
23 encoding lysozyme which resides on the plasmid pLysS (Novagen). T7 lysozyme cuts a
24 specific bond in the peptidoglycan cell wall of *E. coli*. High levels of expression of T7
25 lysozyme can be tolerated by *E. coli* since the protein is unable to pass through the inner
26 membrane to reach the peptidoglycan cell wall. Mild lytic treatments of cells expressing T7
27 lysozyme that disrupt the inner membrane results in the rapid lysis of these cells. Thus, use
28 of the pLysS plasmid should facilitate the lysis of *E. coli* host cells expressing the library
29 polypeptide constructs.

30 *Arraying single-cell clones*

31 Prior to induction of fusion polypeptides, individual clones are arrayed at unique
32 locations. The location from which each library polypeptide is derived will serve to identify
33 it during subsequent screening steps. Each unique location is tracked throughout the

1 screening, either by directly moving each segregated library polypeptide sequentially to
2 other, correspondingly unique locations, or by indirectly tracking the origin of each library
3 polypeptide via its corresponding location-determinable support, which is adhered to the
4 library polypeptide via the adhesion moiety that was incorporated in the above-described
5 fusion construct.

6 Methods for generating single-cell clones are known to the art. For example, the
7 library is first plated to permit well-isolated colonies to grow. Cells from individual colonies
8 may be isolated manually or via automated techniques such a colony picker, and cells from
9 each isolated colony are placed at its corresponding unique location to generate a single-cell
10 clone. Commercially available microtiter trays, for example in 96 or 384 well formats,
11 provide convenient arrays for generating and tracking a unique location for each such single-
12 cell clone. Alternatively, as described in more detail below, the process may be automated
13 for generating arrays with large numbers of single cell-type clones, each of which generates a
14 correspondingly unique library polypeptide.

15 *Lysing the host cells*

16 Following induction and expression, the host cells are harvested and lysed and the
17 polypeptide-bearing lysate collected. A variety of lysing techniques are suitable for use in
18 this invention, including without limitation the three techniques described in detail below.
19 The cells also may be sonicated, for example with the use of commercially available
20 sonicators designed for use with, e.g., 96 well plates (e.g., Misonix Incorporated. Model 431-
21 T).

22 In one embodiment, host cells are lysed using osmotic shock. This technique is a simple
23 method of preparing the periplasmic fraction of expressed proteins. In *E. coli* strains
24 containing the pLysS plasmid, standard osmotic shock techniques can be modified as
25 follows: T7 lysozyme-containing host cells are resuspended in ice-cold 20% sucrose, 2.5mM
26 EDTA, 50mM Tris-HCl pH 8.0 to a concentration of OD₅₅₀=5 and incubated on ice for 10
27 minutes. The cells are centrifuged at 15,000xg for 30 seconds, the supernatant discarded, and
28 the pellet resuspended in the same volume of ice-cold 2.5mM EDTA, 20mM Tris-HCl pH 8.0
29 and incubated on ice for 10 minutes. The cells are centrifuged at 15,000xg for 10 minutes.
30 The supernatant contains protein fraction released due to osmotic shock. Total protein is
31 assessed using the BCA Protein Assay kit.

32 In another embodiment, the host cells are lysed by employing a freeze/thaw protocol.
33 This technique is intended for cells containing the pLysS plasmid. Such cells are

1 resuspended in 1/10 culture volume of 50mM Tris-HCl pH 8.0, 2.5mM EDTA. The cells are
2 frozen at -80°C and then rapidly thawed in order to lyse the cells. The cell debris are pelleted
3 at 15,000xg for 10 minutes and the supernatant saved. To shear the DNA, a DNA nuclease
4 solution is added and incubated for 15-30 minutes at 30°C. The number of freeze/thaw cycles
5 required is determined by monitoring lysate protein concentration.

In yet another embodiment, the host cells are lysed by addition of a mild detergent. This technique is also intended for cells containing the pLysS plasmid. Host cells lacking the pLysS plasmid were resuspended in 1/10 culture volume of 50mM Tris-HCl pH 8.0, 2.0 mM EDTA and 100 µg/ml lysozyme. Cells were then incubated for 15 minutes at 30°C. Triton X-100 was added to a final concentration of 0.1% and incubated for 15 minutes at room temperature. The cell debris were pelleted at 15,000xg for 10 minutes and the supernatant saved. To shear the DNA, a DNA nuclease solution is added and incubated for 15-30 minutes at 30°C.

EXAMPLE 2 PREPARATION OF MICROBEADS

A variety of supports can be used as randomizable supports for binding ligands, and location-determinable supports for binding the library polypeptides. Suitable supports include beads in a variety of sizes and compositions. Selection of a particular bead depends in part upon the type of adhesion to be used (i.e., chemical/covalent linking, or linking through biological adhesion moieties), and the size and type of library polypeptide or other ligand to be adhered to the bead.

One preferred system uses polystyrene microparticles of, e.g., 10 μ m, to adsorb proteins onto the surface of the bead (Polysciences, Inc. or Bangs Laboratories, Inc.). Library polypeptides are adhered to such supports by hydrophobic interactions between the library polypeptides and the bead surface. Other ligands are adhered by, e.g., synthesizing the combinatorial ligand library on the surface of the bead itself, or by incorporating a reactive functional group into the ligand structure, by which a covalent link is formed to the bead surface.

30 The polystyrene beads are exposed to, e.g., the individual library polypeptides
31 uniquely located in the library arrays by suspending an aliquot of the beads in a buffer that is
32 compatible with the chosen lysate solution (e.g., for mild detergent lysis, 1% Triton X-100
33 may be used) and pipetting aliquots into each 384 well format microtiter well. The beads are
34 mixed by repetitive pipetting or by shaking the array plates to ensure maximal dispersion.

1 The beads are left in for approximately 5-15 minutes to several hours, depending on the scope
2 of the population to be screened, to ensure greater than approximately 70-100% maximal
3 adhesion of the polypeptides to the microsupports. Exact conditions are optimized by routine
4 testing familiar to one of ordinary skill in the art. The beads bearing the library polypeptides
5 or other ligands then are removed, for example by vacuuming the soluble contents of each
6 well through the base of a 384 well filter plate and then collecting the remaining coated
7 beads, which are then utilized for interaction screening, as described below.

8 Another preferred embodiment utilizes streptavidin coated polystyrene beads to bind
9 fusion proteins containing biotin. Such beads feature streptavidin molecules saturated to 1.8
10 mgs per gram of 10 μ m polystyrene particle. To form such beads, streptavidin molecules
11 (Pierce) are coupled to polystyrene beads having surface carboxyl reactive groups
12 (Polysciences, Inc. or Bangs Laboratories, Inc.) using techniques familiar to those in the art.
13 The particles are placed in the buffer 2-[N-morpholino]ethanesulfonic acid (MES). They are
14 reacted with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (Pierce)
15 and N-hydroxysuccinimide (NHS) (Pierce) to form an acyl amino ester. Alternatively, the
16 particles are reacted with EDC to form an amine-reactive O-acylurea intermediate, which can
17 then react with the free amine on a polypeptide to covalently link the polypeptide (e.g.,
18 streptavidin) to the bead surface. After washing with MES to remove excess reagents, excess
19 streptavidin (e.g., 18 mgs per gram of bead) is added and the reaction mixed. The derivatized
20 beads are then ready to bind biotin-bearing fusion polypeptides.

21 Still another preferred embodiment utilizes a calmodulin surface coating to bind
22 fusion polypeptide constructs that include the calmodulin binding peptide (CBP). Such beads
23 feature approximately 2.3 mgs calmodulin (Sigma) per 1 gram bead, covalently coupled to a
24 10 μ m polystyrene particle, via the same chemistry described above for covalently linking
25 streptavidin. In embodiments utilizing calmodulin and calmodulin binding protein (CBP),
26 the moieties may be crosslinked as follows. A streptavidin coated 10um particle (prepared as
27 described above) was placed into a bacterial lysate in which a biotin-thioredoxin-CBP
28 (biotrxCBP) fusion protein had been expressed, and the moieties allowed to bind. The beads
29 were then washed to remove nonspecifically bound proteins, and reacted with a commercially
30 available purified calmodulin having FITC covalently attached to the protein (Sigma). In the
31 presence of calcium this interaction takes place (Figure 7). Upon the removal of calcium, the
32 calmodulin/CBP interactions begin to dissociate. However, when the CBP-calmodulin was

1 reacted with a crosslinker such as disuccinimidyl suberate, the calmodulin/CBP interaction
2 remained stable even in the absence of calcium.

3 *Magnetic beads.*

4 In some embodiments, magnetic beads may be used to facilitate collection of the
5 adhered polypeptides, ligands, or interacting pairs. One preferred embodiment of such a
6 magnetic bead features a magnetic core with a polystyrene exterior coating, sized from 1-10
7 μm (commercially available Polysciences, Inc. or Bangs Laboratories, Inc.). Such magnetic
8 beads will bind proteins by direct adsorption, via the polystyrene coating. Alternatively,
9 streptavidin-coated magnetic beads may be used. A variety of sizes are suitable, including
10 135 nm diameter beads (Immunicon, Inc.), 50nm diameter beads (Miltenyi Biotec Inc.), 1 μm
11 diameter beads (Bangs Laboratories), 2.8 μm diameter beads (Dynal Inc.) and 5 μm diameter
12 beads (CPG Inc.). In still another embodiment, calmodulin coated magnetic particles are
13 used. Such particles are synthesized by the same technique described above for streptavidin
14 coated microparticles, but with the exception that calmodulin is substituted for streptavidin.
15 Again, the starting particle is a magnetic particle with carboxy functional groups on the
16 surface (Bangs Laboratories or Polysciences, Inc.).

17 Interactions of a protein on a 10 μm polystyrene bead and a protein on a 150nm
18 magnetic bead were carried out in two systems. In one system, one set of 10 μm beads
19 (prepared as described above) were coated with biotin, and another set of 150nm magnetic
20 beads (Immunicon) were coated with streptavidin. A reaction tube was set up with 10^6 BSA
21 coated 10 μm beads, about 200 10 μm biotin coated beads and about 10^8 150nm streptavidin
22 coated particles in PBS with 0.5% BSA. Figure 8. These were reacted together for fifteen
23 minutes to allow for binding between the biotin and streptavidin moieties. In order to enrich
24 for these aggregates, a neodymium-iron-boron magnet was placed to the side of the tube and
25 the liquid removed. After several washes with PBS the number of biotin coated and BSA
26 coated particles were counted with a hemacytometer. It was found that the mixture had been
27 enriched several thousand fold for the biotin coated particles.

28 The other system examined the interaction of SV40 large T antigen with an antibody
29 to the antigen. First, streptavidin coated 10 μm beads prepared as described elsewhere herein
30 were added to a lysate containing a biotin thioredoxin SV40 large T antigen fusion protein
31 (prepared as described elsewhere herein). About 200 of these large T antigen coated beads
32 were added to a mixture of about 10^6 BSA coated 10 μm beads, along with some 10^{10} 150nm
33 magnetic beads coated with goat anti-mouse secondary antibodies (Immunicon), and 0.5ug of

1 mouse anti-SV40 large T antigen (Santa Cruz). Figure 9. The reaction proceeded as above.
2 Again the enrichment was several thousand fold for the 10um SV40 large T antigen coated
3 beads.

4 *Fluorescence-tagged beads.*

5 In order to distinguish one type of ligand from another, each such ligand may be
6 adhered to a randomizable support that bears a corresponding unique tag. One way of
7 creating a unique tag is to adhere to the exterior surface of a nonporous randomizable
8 support, or to entrap within interior regions of a porous randomizable support, a particular
9 mixture of fluorescent dyes - a unique fluorescent dye identifies, also referred to herein as a
10 fluorescent "bar code". The fluorescent dyes may be organic in nature, or alternatively may
11 be fluorescent nanoparticles. Two variables contribute to the bar code -- type of dye (i.e., its
12 particular emission spectrum) and concentration of dye (i.e., intensity of its emission signal).
13 A wide variety of fluorescent dyes with well-characterized excitation and emission spectra
14 are commercially available. For example, Molecular Probes, Inc provides a variety of
15 organic dyes; (see TABLE 1, below). Alternatively, fluorescent nanoparticles may be
16 obtained that feature specific excitation and emission spectra. Such nanoparticles are
17 described by Bruchez et al, *Semiconductor nanocrystals as fluorescent biological labels*,
18 *Science* 281: (5385):2013-16 (Sept. 1998) and Cahn, W.C. and Nie, S, *Quantum dot*
19 *bioconjugates for ultrasensitive nonisotopic detection*, *Science* 281(5385):2016-18 (Sept.
20 1998), the disclosures of which are incorporated herein in their entireties. Indeed, it is
21 possible to procure sets of fluorescent molecules that cover the spectrum from blue to red.
22 Each dye has characteristic excitation and emission spectra that may be used to create a bar
23 code.

24 In one embodiment of the invention, a set of fluorescent bar codes is created that is
25 sufficiently large to uniquely identify each member of a ligand pool on the order of 1×10^6
26 members (i.e., roughly each protein encoded by a human cell). Optimally, the corresponding
27 set of unique tags is generated from a set of 4-10 separate fluorescent dyes. The dyes are
28 chosen so that there is optimal compatibility of their excitation and/or emission maxima when
29 such dyes are irradiated by any one of a given FACS machine lasers, including Argon and
30 Helium-Neon. The dyes are selected further so that there is minimal overlap of their
31 emission maxima. Moreover, the dyes are chosen so as to be distinguished from any
32 autofluorescence emissions of the bead to be labeled. However, as described below, it is
33 possible to choose dyes that have some overlap because the dye cross-talk can be

- mathematically reduced or eliminated by certain computations that can be performed off line (i.e., by computers that use stored fluorescence data files as input).

TABLE 1
EXEMPLARY ORGANIC DYE SPECIES

Molecular Probes, Inc. Catalog #	Dye Name	Excitation wavelength (nm)	Emission maxima (nm)
A-191	7-amino-4-methylcoumarin	351	430
B-3932	bodipy® 665/676	665	676
C-652	5-(and-6)-carboxynaphtho	599	667
D-113	dansyl cadaverine	335	520
D-275	DiOC18	484	499
D-282	DiOC18(3)	548	564
D-307	DiOC18(5) oil	644	663
D-2184	Biodipy® FL, SE	488	530
D-2186	bodipy® 530/550	530	550
D-2187	bodipy® 530/550SE	530	550
D-2190	bodipy® 493/503	493	503
D-2191	bodipy® 493/503SE	493	503
D-2219	Bodipy® 558/568, SE	558	568
D-2221	bodipy® 561/570	561	570
D-2222	bodipy® 564/570SE	564	570
D-2225	Bodipy® 576/589	576	589
D-2227	bodipy® 581/591	581	591
D-2228	Bodipy® 581/591,SE	581	591
D-3921	bodipy® 505/515	505	515
D-3922	bodipy® 493/503	493	503
D-6102	Biodipy® FX-X, SE	488	530
D-6117	Bodipy® TMR-X,SE	540	560
D-6180	Bodipy® RGG,SE	530--	550
D-6186	Bodipy® R6G-X, SE	530	550
	fluorescein		
D-10000	Bodipy® 630/650-X,SE	630	650
D-10001	Bodipy®650/665-X,SE	650	665
D-12731	DiOC18(7)	748	780
N-1142	nile red	552	636

In other embodiments, fluorescent nanocrystals (Quantum Dot, Corp., Palo Alto CA) may be utilized as the fluorescent dye species forming the barcode. Briefly, the nanocrystal is a semiconductor material such as zinc sulfide-capped cadmium selenide. The nanocrystal also may feature an outer layer to aid in derivatization and/or to aid solubility, for example mercaptoacetic acid (Chan and Nie (1998), *supra*), or silica derivatives (Bruchez et al.

1 (1998), *supra*. The emission spectrum of the nanocrystal is dependent upon the size of the
2 cadmium selenide core of the crystal.

3 Fluorescent nanocrystals may be coupled with the beads in a variety of ways. One
4 general approach is to apply absorption techniques such as are used in absorbing organic
5 fluorochromes to beads. Briefly, the nanocrystals can be rendered nonpolar for this purpose
6 by coating the nanocrystals with a nonpolar coating such as an alkyl silane. A polystyrene
7 bead having a porous structure is then exposed to the nonpolar fluorescent nanocrystals, using
8 methods familiar to those in the art. The nanocrystal then equilibrates into the corresponding
9 nonpolar interior of the polystyrene bead, and is maintained there by repulsion from an
10 aqueous solvent. Optionally, more porous particles (Dyno Particles, Inc.) may be utilized to
11 increase the available interior region.

12 Alternatively, the nanocrystals may be linked to the selected beads via covalent
13 bonds, using a variety of different chemistries familiar to those of skill in the art. In such
14 embodiments both the bead surface and the nanocrystals are derivatized with surface reactive
15 groups. In some embodiments, the bead features a porous surface, allowing the nanocrystals
16 to diffuse into the interior regions of the bead prior to covalently cross-linking with the bead.
17 In other embodiments, nonporous bead particles may be used, in which case the nanocrystal
18 is crosslinked to the exterior surface of the bead.

19 A variety of beads and crosslinking chemistries are suitable for use in this invention.
20 For example, in some instances it is advantageous to use porous silica particles having low
21 autofluorescence. As one nonlimiting example, carboxyl coated silica particles (CPG, Inc.)
22 of a desired size (e.g., 10 μm diameter) are selected. The nanocrystals are first reacted with
23 an amine silane, thereby forming an amine functional group. The derivatized beads and
24 nanocrystals are then mixed together so that the nanocrystals diffuse evenly throughout the
25 particle. A crosslinking agent such as EDC (1-ethyl-3-(3-
26 dimethylaminopropyl)carbodiimide) is then added, thereby conjugating the nanocrystal to the
27 derivatized silica particle. In other embodiments, other derivatized particles may readily be
28 substituted.

1 *Fluorescence barcoding.*

2 The barcoding system uses a set of dye species chosen with the considerations
3 enumerated above, as exemplified but not limited to those dyes in Table 1 or the nanocrystals
4 described above. The identity of each randomizable support is encoded as a numerical
5 readout having digit placeholders equal to the number of dyes used (e.g., nine dyes create
6 nine "digits" in the barcode). Each digit in the barcode is then further defined by the amount
7 of the specific dye, as determined from its fluorescence intensity (i.e., 0x, 1x, 2x, 3x or 4x).
8 Thus, for 9 dyes and 5 amounts (or fluorescence levels) there are $(5)^9$ possible barcodes.

9 The beads are labeled with dyes by mixing the selected number of dyes in defined
10 ratios such that a specific bead receives a unique barcode. For example, using nine different
11 dyes one defined bead type may receive dyes in the ratio of (4, 2, 3, 3, 1, 1, 2, 4, 2); a second
12 bead type may receive dyes in the ratio (2, 2, 3, 3, 1, 1, 2, 4, 2). These beads differ only in
13 the levels of the first dye (the first bead type has level 4, the second has level 2).

14 Fluorescent organic dye species may be selected from a wide variety of known dyes
15 and incorporated into a wide variety of known beads, utilizing techniques familiar to those of
16 skill in the art. E.g., U.S. Patent No. 5,573,909, the disclosure of which is incorporated by
17 reference herein in its entirety. As a non-limiting example, by mixing the dyes in an organic
18 solvent such as, e.g., acetonitrile or dimethylformamide, and adding the dye solutions in
19 defined ratios to individual groups of beads and allowing the absorption reactions to go to
20 completion, it is possible to irreversibly adsorb dye molecules onto the bead surface and
21 interior. Removal of the organic solvent followed by drying, leaves the beads labeled with
22 the nine dyes in the predetermined amount dispersed over the surface of each bead.
23 Fluorescently labeled beads prepared in this general way but with only one or a few
24 fluorescent tags have been described in the literature (Michael et al., *Analytical Chemistry*
25 70(7):1242-48 (1998); Fulton et al., *Clinical Chemistry* 43(9):1749-56 (1997)) and are
26 available commercially (Luminex Corp.).

27 As one non-limiting example of the barcoding strategy, four dyes were selected for
28 study: BioDIPY 493N, BioDIPY 560PA, BioDIPY580PA and BioDIPY665N. The dyes
29 were incorporated into polystyrene beads (Bangs Labs, Inc. PS07N) beads as follows. The
30 selected dyes were dissolved in dimethylformamide (DMF). The beads were washed three
31 times with absolute ethyl alcohol (and stored in same). A staining mix was prepared,
32 containing 10% DMF, 54% absolute ethyl alcohol and 36% dichloromethane (approximating
33 a 60:40 ratio of ethyl alcohol to dichloromethane). The beads were added and rapidly stirred

for ten minutes. The staining solution was then removed from the beads by centrifugation or filtration and the beads were washed two times with absolute methanol followed by two washes of PBS/1% TWEEN 20. The dyed beads were then stored in the PBS/TWEEN 20 mixture at 4° C, protected from light. The beads were doped with five different concentrations of each dye, as summarized below in Table 2.

TABLE 2
SUMMARY OF DYE PROFILES

DYE	BARCODE LEVEL	CONCENTRATION (µM)
BIODIPY® D-2190 (NONPOLAR) EX 493 NM / EM 503 NM		
	1	1
	2	0.43
	3	0.1
	4	0.043
	5	0.01
BIODIPY® D-2221 (PROPIONIC ACID) EX 460 NM / EM 570 NM		
	1	159
	2	68
	3	16
	4	6.8
	5	1.6
BIODIPY® D-2227 (PROPIONIC ACID) EX 580 NM / EM 590 NM		
	1	132
	2	57
	3	13
	4	5.7
	5	1.3
BIODIPY® B-3932 (NONPOLAR) EX 665 NM / EM 676 NM		
	1	100
	2	43
	3	10
	4	4.3
	5	1

Next, the fluorescence intensity of each dye was characterized in isolation of the others, at five different levels. Table 3 summarizes the resulting fluorescence levels detected in four different windows -- FL1 (525 nm +/- 10 nm), FL2 (575 nm +/- 7 nm), FL3 (620 nm +/- 13 nm) and FL4 (675 nm +/- 15 nm). For each of the four dyes, the fluorescence intensity decreased proportionally to the decreasing dye content of the bead. Moreover, each dye provided a suitably distinct fluorescence signature.

7 Next, the four selected dyes were mixed in varying combinations of dyes/intensity
8 levels, as shown in Table 3. The resulting fluorescence intensities were as shown,
9 demonstrating that the resulting beads provided discernable labeling information regarding
10 both dye concentration and composition.

TABLE 3
FOUR DYE FLUORESCENCE CODING

BODIPY 493N	BODIPY 560PA	BODIPY 580PA	BODIPY 665N	FL1	FL2	FL3	FL4
LEVELS	LEVELS	LEVELS	LEVELS				
1				537	8.5	4.8	1
2				248.4	3.8	2.2	1
3				45	1.2	1.1	1
4				20.4	1.1	1.1	1
5				3.9	1	1	1
	1			43.9	304.2	427.8	17.2
	2			19.4	124.4	180.6	7
	3			3.3	20.7	33.5	1.4
	4			1.5	8.6	13.9	1.1
	5			1.3	2.3	5.3	1
		1		94.9	20.1	345	19.7
		2		37.4	8.2	140.4	7.8
		3		6.3	1.6	30	1.7
		4		2.5	1.2	12.8	1.1
		5		1.2	1.1	3.4	1
			1	4	1.6	11.2	55.8
			2	1.7	1.2	5.3	25.6
			3	1.1	1	2.3	5.3
			4	1	1	1.7	2.3
			5	1	1	1.4	1
1	1			505.4	294.3	446.3	18.8
1		1		529.7	24.8	334.7	19.2
1			1	450.5	7.8	14.5	55.4
	1	1		117.6	299.7	796.1	41.6
	1			41	234.5	361.4	74.9
		1		63	15.7	260.6	74.7

4	1			60.3	268.8	443.8	20.4
4		1		87.5	18	326.4	20.3
4			1	22.2	2.1	11.4	58.2
1	4			505.8	12.8	20.1	1.3
	4	1		75.4	23.5	363.2	22.8
	4		4	4.8	5.5	22.3	60.5
1		4		497.5	8.5	16.2	1.3
	1	4		43.9	266.6	454	21.5
		4	1	5.5	2.2	19.8	59.6
1			4	489.8	7.8	5	2.9
	1		4	41.4	261.7	438.5	22.9
		1	4	72.7	17.6	327.8	22.8

1

2 *Oligonucleotide-tagged beads.*

3 In some embodiments, it is possible to construct a sufficient number of unique
 4 oligonucleotide tags and to attach such tags to the randomizable supports by, e.g., linking the
 5 oligonucleotide to a biotin linker and adhering that linker to a streptavidin-coated bead such
 6 as those described above. The oligonucleotide tags bear unique DNA sequences, each of
 7 which can be correlated to a given ligand.

8 Such DNA tags can be built in one of several ways. For example, using techniques
 9 well known to the art, a multichannel oligonucleotide synthesizer can generate a set of DNA
 10 molecules with unique sequences of any given length. Once individual oligonucleotide tags
 11 characterized and isolated into homogeneous tag pools, the tags can be adhered to the
 12 randomizable supports in a variety of ways. For example, if the randomizable supports have
 13 a streptavidin coating, then a biotin adhesion moiety is joined to each oligonucleotide tag at
 14 the 5' end by standard synthesis techniques. If the randomizable support is coated with other
 15 adhesion moieties, the complementary adhesion moiety can be chemically coupled to a 5'
 16 amino-modified oligonucleotide tag.

17 The oligonucleotide tags may be read either by sequencing, by evaluating sequence
 18 length, or by hybridization. For sequencing information, the oligonucleotide tags resident on
 19 each bead are subjected to PCR, and then run on a sequencing gel. Alternatively, the
 20 oligonucleotide tags may be identified via exposing the tags to known hybridization probes.

21 *Mass spectrometry tags*

22 Another suitable method for encoding identities of beads involves use of mass tags --
 23 i.e., labels that can be detected by mass spectrometry. Such mass tags are known in the art
 24 and must be coupled to the beads in different amounts so as to generate a mass tag bar code.

1 This code can be read by subjecting the beads to mass spectrometry pursuant to methods
2 familiar to those of the art, or by use of gas chromatography.

3 *Radio-frequency tags.*

4 As yet another alternative for encoding identity information on beads, the beads may
5 be engineered to emit unique, identifying radio signals of various pre-determined frequencies.
6 Such beads may contain, e.g., miniaturized transmitter/receiver circuitry, rectifier, control
7 logic and antenna. Each set of beads thus may contain a unique label laser-etched on the
8 internal chip within the bead. Emissions from the radio-frequency tags are detected by a
9 corresponding radio-frequency detector.

10 *Beads with mixed tags.*

11 In some embodiments, the number of different ligand populations to be uniquely
12 tagged will be quite large -- on the order of 1×10^6 or more. Although a corresponding
13 number of unique fluorescent bar code tags, mass spec tags or DNA oligonucleotide tags
14 could be formulated as described above, in some instances it may be desirable to make tags
15 that are some combination of fluorescent, mass spec and/or oligonucleotide information. For
16 example, oligonucleotide tags or mass spec tags may be incorporated so as to reduce the
17 number of fluorescent dyes used. Such techniques may advantageously reduce or avoid any
18 instances of fluorescent quenching or fluorescence resonance energy transfer (FRET), and/or
19 may expand the number of bar codes that can be used.

20 *Bead-polypeptide interactions*

21 To test the bead:bead interactions of the invention, several proteins were inserted into
22 pET-biotrx-BirA and overexpressed in BL21 (DE3) *RIL* cells: murine p53, SV40 large T-
23 antigen, HPV16 E7 and the "Rb pocket" of the Retinoblastoma gene. The E7 and p53
24 polypeptides were bound to the beads via the associated biotinylation signal, and were
25 detected on the beads with antibodies specific to E7 and p53, respectively.

26
27 **EXAMPLE 3**

28 **HIGH THROUGHPUT SCREENING OF A COMPREHENSIVE HUMAN PROTEIN**
29 **LIBRARY FOR PROTEIN-PROTEIN INTERACTIONS**

30
31 The goal of the process is to examine in a quantitative or semi-quantitative fashion all
32 possible pairwise interactions between human protein domains. This involves a test of " $n \times$
33 n" interactions, if " n " is the number of human protein domains. Values for " n " likely fall

1 between 100,000 and 1,000,000. For an interaction screen of this scope, automation of at
2 least some of the following procedures is desirable.

3 To summarize, one embodiment of the process involves a series of steps: (1)
4 generation of a library of expressed human sequences in an *E. coli* expression vector such
5 that the human DNA is expressed as a fusion with a suitable adhesion moiety; in addition,
6 part of the fusion protein may serve as a recognition sequence tag for attaching labels (e.g.,
7 fluorescent antibody labels) so that the protein can be detected; (2) enrichment of the library
8 for clones that contain constructs that are in-frame and expressed at reasonable levels; (3)
9 arraying of the enriched library clones in microtiter plates; (4) growth and induction of the
10 individual library clones to produce fusion proteins inside *E. coli*; (5) preparation of *E. coli*
11 lysates to release the expressed fusion proteins from cells; (6) generation of a primary set of
12 beads barcoded with suitable combinations of fluorescent dyes to act as randomizable
13 supports; (7) apportioning of beads to individual wells of microtiter trays to permit adhesion
14 of lysate fusion proteins to the randomizable supports (also referred to herein as "primary
15 beads"); (8) apportionment of secondary magnetic beads (as location-determinable supports)
16 to microtiter wells to allow adhesion of lysate proteins as in 7; (9) mixing of primary and
17 secondary beads to permit aggregation of beads with interacting proteins on their surfaces;
18 (10) magnetic capture of secondary beads and attached primary beads to enrich for primary
19 beads with proteins that interact with protein on the surface of secondary beads; (11) mixing
20 of enriched primary beads with soluble fusion protein in microtiter wells to allow interaction
21 of soluble protein with proteins on the surface of primary beads, as well as detachment of
22 secondary beads; (13) magnetic capture and disposal of secondary beads; (14) collection of
23 primary beads and crosslinking of bound protein using, e.g., paraformaldehyde; (15)
24 exposure to labeling agent (e.g., fluorescent antibody) to enable detection of bound secondary
25 proteins; and (16) detection of labeling agent and barcode reading to determine identity of
26 primary protein (on bead surface) and amount of secondary protein attached via interaction
27 with primary protein. Other embodiments may add to, alter or delete some of the above
28 steps, in ways that will be apparent to one of ordinary skill in the art.

29 Steps 1 and 2 --generation and enrichment of the polypeptide library to be cross-
30 screened in order to generate a protein interaction may -- is described in detail in Example 1,
31 above.

32 Step 3 involves plating out and growing up single-cell clones that produce only one of
33 the library polypeptides at a given unique array location. To accomplish this, a commercial

1 robot may be used (e.g., Genetix Ltd. "Q-bot™; TM Analytic, PBA Flexys™; BioRobotics
2 Ltd, BioPick™; or Linear Drives Ltd., Mantis™; any of which with multiple pin tool picking
3 head) to select out a single colony and transfer the cells to a corresponding unique array
4 location in e.g., a 384 well microtiter plate (40 µl volume). Each clone in the array is grown
5 in, e.g., Luria broth or minimal media until early- to mid- log phase, and then expression of
6 the human protein domain construct is induced by adding IPTG (step 4). After a suitable
7 period of time to allow polypeptide expression, the cells are then lysed (step 5) by the method
8 described in detail in Example 1. Thus, each unique location in the chosen array format (384
9 well plate or other) will contain a lysate bearing one particular human protein domain,
10 amongst the milieu of native *E. coli* proteins.

11 Alternatively, for ease of generating and processing the lysate from the single cell
12 clone, a single colony may be picked and transferred to a correspondingly unique
13 intermediate container of larger volume for growing up the clone. Once the clone is finished
14 culturing, a sample is taken from the intermediate container and is concentrated and lysed as
15 described in detail in Example 1. An aliquot of the lysate is then transferred to a unique array
16 location in a 384 well microtiter plate (40 µl volume).

17 Step 6 involves the generation of the primary set of beads with fluorescent barcodes.
18 These beads are the randomizable supports that will allow presentation of an aliquot bearing a
19 fully integrated collection of lysate protein domains to each such domain independently, to
20 map all possible interactions amongst those protein domains. Example 2 describes
21 preparation of these uniquely tagged fluorescent beads in detail.

22 Once each primary set of beads with a corresponding unique fluorescent tag is
23 generated, the bead sets are suspended in buffer. A sampling from each tagged bead set is
24 then dispersed into a corresponding array location, so that the tagged primary beads adhere to
25 the protein domains therein (step 7). This may be accomplished by e.g., automated aspiration
26 of the beads into the wells (e.g., TecanAG Genesis™; Matrix Technologies Corp.
27 PlateMate™; Carl Creative Systems, Inc. PlateTrak™) or hopper release of beads into wells.
28 Conversely, an aliquot of the lysate may be aspirated or released from a hopper into a
29 corresponding microtiter well that already contains these primary fluorescent beads. In either
30 event, the beads and protein domains are brought into contact and allowed to adhere via the
31 adhesion moiety fused to the protein domain. The identity of the adhered protein thereafter
32 can be determined via the corresponding, unique fluorescent bar code tag on the bead.

Once each of the unique array locations (i.e., polypeptides or other substrates) has been exposed to a corresponding set of beads bearing a unique tag, all beads are collected and mixed to form a fully integrated set of protein-bearing beads. This random mixing is accomplished by multiple, automated aspiration and release cycles, by plate agitation with a robotic shaker, or by mechanical stirring.

Next, the secondary set of magnetic beads are prepared *in situ* in each of the unique locations in the library array (step 8). This is accomplished by adding an aliquot of beads to each library as in step 7. Alternatively, a robotic hand with magnetized fingers may be used to capture the magnetic beads and then release the beads in each of the corresponding array locations on the, e.g., 384 well plate, by dipping the fingers into the lysate and demagnetizing the fingers.

Aliquots taken from the fully integrated set of primary beads are then collected and dispensed into each unique array location, each of which contains a location-determinable set of secondary beads with adhered protein domains (step 9). The number of primary beads (i.e. randomizable substrates) should be sufficient to reduce probability of *not* having a particular polypeptide/bead to a small value -- e.g., less than 1:100 probability. This may be accomplished by aspirating and dispensing, as above. This step allows complexes to form between the protein domains adhered to the primary and secondary beads at each array location, and hence forming bead-bead aggregates.

Complexes of adhered beads are then retrieved magnetically (step 10) with, e.g., a neodymium-iron-boron magnet (Master Magnetics Inc.). The magnetic aggregates using relatively large magnetic beads (i.e. larger than about 50 nm diameter) are magnetically attracted to the sides of the microtiter wells, either on one side or around the entire perimeter of the wells. Remaining beads are washed away. As yet another alternative, a ferromagnetic pin is placed in the center of the well, with magnets located on the outside of the well. Geometry of the pin and magnet is selected so that the induced magnetic field on the pin will attract the beads, and beads that do not react are removed.

Quantification of polypeptide-ligand complexes may be facilitated by replacement of bead-bound protein domain with a soluble, unbound form of the domain (step 11). This is accomplished by introducing the enriched bead complexes derived from step 10 into a soluble protein domain lysate that matches the protein domain on the secondary bead (i.e., the location-determinable domain). Alternatively, the beads may be exposed to the products of a separate library that contains polypeptide inserts that correspond to each polypeptide moiety

1 that is adhered to the bead, but which has a unique labeling domain or epitope. This is
2 readily accomplished by placing the complexes that correspond to, e.g., an array location
3 designated "1" in a first a set of primary 384-well microtiter trays (step 3) into a
4 corresponding location, e.g., designated "1'", of a duplicate microtiter tray that was prepared
5 in parallel in step 3. Since array location 1 and 1' contain the same lysate, the free lysate in
6 1' will competitively displace the bead-bound lysate of the complex. As a result, the primary
7 bead will now bear two layers of protein domains, adhered to one another via protein-protein
8 interactions.

9 Once the protein-protein interactions are established, the primary beads are collected
10 in a manner that segregates the beads in groups that correspond to each separate array
11 location from which the protein bound to the secondary bead originated and the bound
12 proteins crosslinked with, e.g., paraformaldehyde (step 14) to stabilize the complexes by
13 preventing dissociation.

14 These stabilized protein-protein pairs are then exposed to a fluorescent antibody (step
15). As one non-limiting example, one may detect a bound secondary protein by using a
16 fluorescently-labeled antibody directed against one of the fusion protein epitopes (used as a
17 recognition domain and shared among all library constructs), e.g., a FLAG or biotin epitope.
18 The antibody is incubated with the crosslinked beads, such that it binds to exposed or unique
19 epitopes on the secondary protein; i.e., the labeling agent must recognize an epitope that is
20 either absent from the primary fusion polypeptide, thus necessitating construction and array
21 of a separate library for the secondary polypeptide, or an epitope that is inaccessible on the
22 primary polypeptide). Alternatively, fluorescently labeled avidin may be used. These beads
23 are washed in binding buffer and then analyzed as described below. The fluorescence
24 intensity of the antibody fluorochrome serves as a surrogate for the amount of bound
25 secondary protein.

26 Finally, in step 16, the beads bearing these segregated, labeled protein pairs are then
27 examined by a detecting device to quantify conjugates that have the antibody or biotin label.
28 In one preferred embodiment, the fluorescence information (both wavelength and intensity
29 signatures) are simultaneously read and used to identify the protein domain adhered to that
30 bead. Alternatively or in conjunction, the beads are decoded using familiar techniques such
31 as sequencing or hybridization of oligonucleotide tags, or mass spectrometry to identify mass
32 tags.

1 This sorting and/or detection step can be accomplished via one of a number of
2 instruments. Two general categories of instrument have particular utility: a flow cytometry
3 instrument such as a FACS machine or flow analyzer; CCD detector or photomultiplier tube
4 scanner. Each device must have certain capabilities. It must permit rapid analysis of beads
5 using, in the case of FACS, multiple lasers for excitation (e.g., three lasers), and detection of
6 fluorescent emissions at multiple wavelengths (e.g., 3-10 wavelengths). Such capabilities
7 presently exist in the Cytomation flow sorter. The three lasers excite cells or beads in liquid
8 droplets sequentially as the droplets fall in a stream. A series of filters and photo-multiplier
9 tubes (PMTs) then collect emitted light at different preselected wavelengths. These data are
10 stored and can be accessed for analysis later off-line from files.

11 The bead barcode reveals the identity of the primary protein by correlating that
12 protein back to a unique library array location --i.e. the microtiter well that contained the one
13 particular lysate that was exposed to that barcoded primary bead. This barcode is read in the
14 same step as the antibody quantitation is performed. However, to decode a large number of
15 bar codes, multiple measurements on each bead are required. For example, it may be
16 necessary to measure fluorescence emissions of ten dyes at ten wavelengths with specific
17 excitation lasers. These ten measurements provide sufficient information to unambiguously
18 identify each bead according to its specific barcode.

19 The process by which this computation is performed involves two basic steps: (1)
20 parameters are fit to known barcode data; (2) the fitted parameters are used in a
21 deconvolution calculation to determine the bar codes of unknown beads. Total fluorescence
22 of a barcoded bead at a particular wavelength (and at a particular excitation wavelength) can
23 be calculated according to a formula:

24 $F = l_1 f_1 + l_2 f_2 + \dots + l_n f_n$

25 where l_i is the quantity or level of the first dye and f_i is the normalized fluorescence
26 contribution of the first dye under particular conditions of excitation and emission (i.e.,
27 wavelengths). By generating many beads with defined dye ratios (i.e., bar codes) and
28 measuring their fluorescence (F) at specific wavelengths, it is possible to fit the f_n parameters
29 and create at specific wavelengths a set of equations that relate total fluorescence to the
30 individual fluorescences of the different dyes. After this is completed, it is possible to
31 calculate the l_n 's of an unknown bead, thereby determining its barcode and identity. It is
32 necessary to have at least as many independent measurements of F (i.e., at different
33 excitation/emission wavelengths) as there are unknown "l" values in the bar code.

1 The fluorescent barcode is used to determine the bead identity, an identity that is
2 linked to the well from which it was originally derived; that is, a barcode matches a well
3 which contained the lysate fusion protein that comprises layer one on the bead. Thus, the
4 nature of the first layer of protein that is adhered to the support can be determined by DNA
5 sequence analysis of the cloned insert in each well. This sequence analysis can be
6 accomplished simply by PCR amplification of insert sequences from each microtiter well
7 using primers on the vector which flank the insert. Standard automated sequence analysis
8 followed by database searches reveals details about each cloned insert. Current sequencing
9 throughputs permit sequencing of one million inserts in a period of weeks to months.

10 As described above, the fluorescence of a labeling agent, e.g., an antibody against a
11 FLAG epitope serves to quantify the amount of secondary protein attached via protein-
12 protein interactions to a bead. If the concentration of protein in the lysate is measured or
13 estimated, and the saturating amount of protein on the bead is known (i.e., how much
14 secondary protein could be maximally bound if all primary protein binding sites were
15 occupied), it is possible to determine the approximate binding constant of the protein-protein
16 interaction from the equation:

17 $K_d = [xy]/[x][y]$

18 where the ratio $[xy] / [x]$ is simply the ratio of measured bound secondary protein over the
19 saturating (maximal) bound amount, and $[y]$ is concentration of soluble fusion protein in the
20 lysate.

21

* * *

22 While the present invention has been described in terms of specific methods and
23 compositions, it is understood that variations and modifications will occur to those skilled in
24 the art in consideration of the present invention. Accordingly, it is intended in the appended
25 claims to cover all such equivalent variations which come within the scope of the invention
26 as claimed, in light of those variations and modifications.

27

CLAIMS

- 1
- 2
- 3
- 4 1. A method for identifying interacting substrate-ligand pairs, comprising the steps of:
- 5 (a) adhering a plurality of ligands to a corresponding plurality of randomizable
- 6 supports bearing a unique fluorescent dye identifier;
- 7 (b) contacting said ligands with a substrate derived from a unique location so as to
- 8 form at least one substrate/ligand complex;
- 9 (c) identifying any complex-forming ligand by its corresponding unique
- 10 fluorescent dye identifier; and
- 11 (d) identifying any complex-forming substrate by determining its corresponding
- 12 unique location.
- 13
- 14 2. The method of claim 1, wherein said substrate is an individual polypeptide.
- 15
- 16 3. The method of claim 1, wherein said substrate is a library polypeptide.
- 17
- 18 4. The method of claim 3, wherein said library polypeptide is a native polypeptide.
- 19
- 20 5. The method of claim 3, wherein said library polypeptide is a member of a large library.
- 21
- 22 6. The method of claim 3, wherein said library polypeptide is a member of a very large
- 23 library.
- 24
- 25 7. The method of claim 3, wherein the identity of said library polypeptide is not known prior
- 26 to step (a).
- 27
- 28 8. The method of claim 1, wherein said ligands are polypeptides.
- 29
- 30 9. The method of claim 8, wherein said ligands are library polypeptides.
- 31
- 32 10. The method of claim 9, wherein said library polypeptides are native polypeptides.
- 33

- 1 11. The method of claim 9, wherein said library polypeptides are members of a large library.
- 2
- 3 12. The method of claim 9, wherein said library polypeptides are members of a very large
- 4 library.
- 5
- 6 13. The method of claim 8, wherein the identities of said polypeptides are not known prior to
- 7 step (a).
- 8
- 9 14. The method of claim 7, wherein each said substrate derived from a unique location is
- 10 adhered to a corresponding location determinable support.
- 11
- 12 15. The method of claim 1 wherein said randomizable support is magnetized and said
- 13 complexes are segregated by being magnetically culled.
- 14
- 15 16. The method of claim 14, wherein said location determinable support is magnetized and
- 16 said complexes are segregated by being magnetically culled.
- 17
- 18 17. The method of claim 15, wherein said randomizable supports are beads.
- 19
- 20 18. The method of claim 1, wherein said unique fluorescent dye identifier is comprises a
- 21 plurality of fluorescent dye species.
- 22
- 23 19. The method of claim 18, wherein said plurality of fluorescent dye species includes at least
- 24 one species of fluorescent nanoparticle.
- 25
- 26 20. The method of claim 18, wherein said plurality of fluorescent dye species includes at least
- 27 one species of organic dye.
- 28
- 29 21. The method of claim 20, wherein said organic dye species is selected from the group
- 30 consisting of the organic dyes listed in Table 1.
- 31
- 32 22. The method of claim 1, wherein said ligands are non-proteinaceous organic molecules.
- 33

1 23. The method of claim 1, wherein said step of identifying comprises the step of detecting
2 each said substrate/ligand complex with a fluorescent label.

3

4 24. The method of claim 22, further comprising the step of detecting said substrate/ligand
5 complex with a CCD camera..

6

7 25. A human protein interaction map produced by the method of claim 1.

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

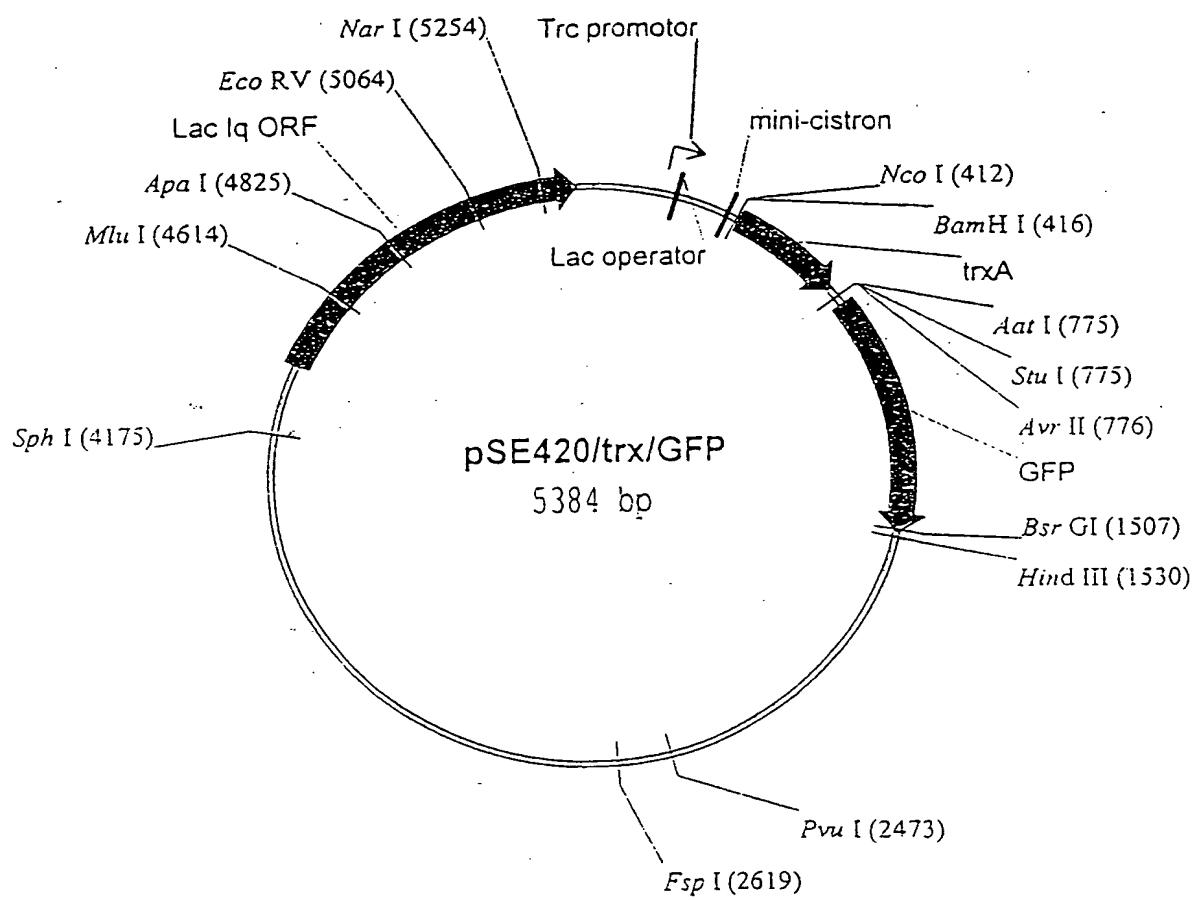


FIGURE 1

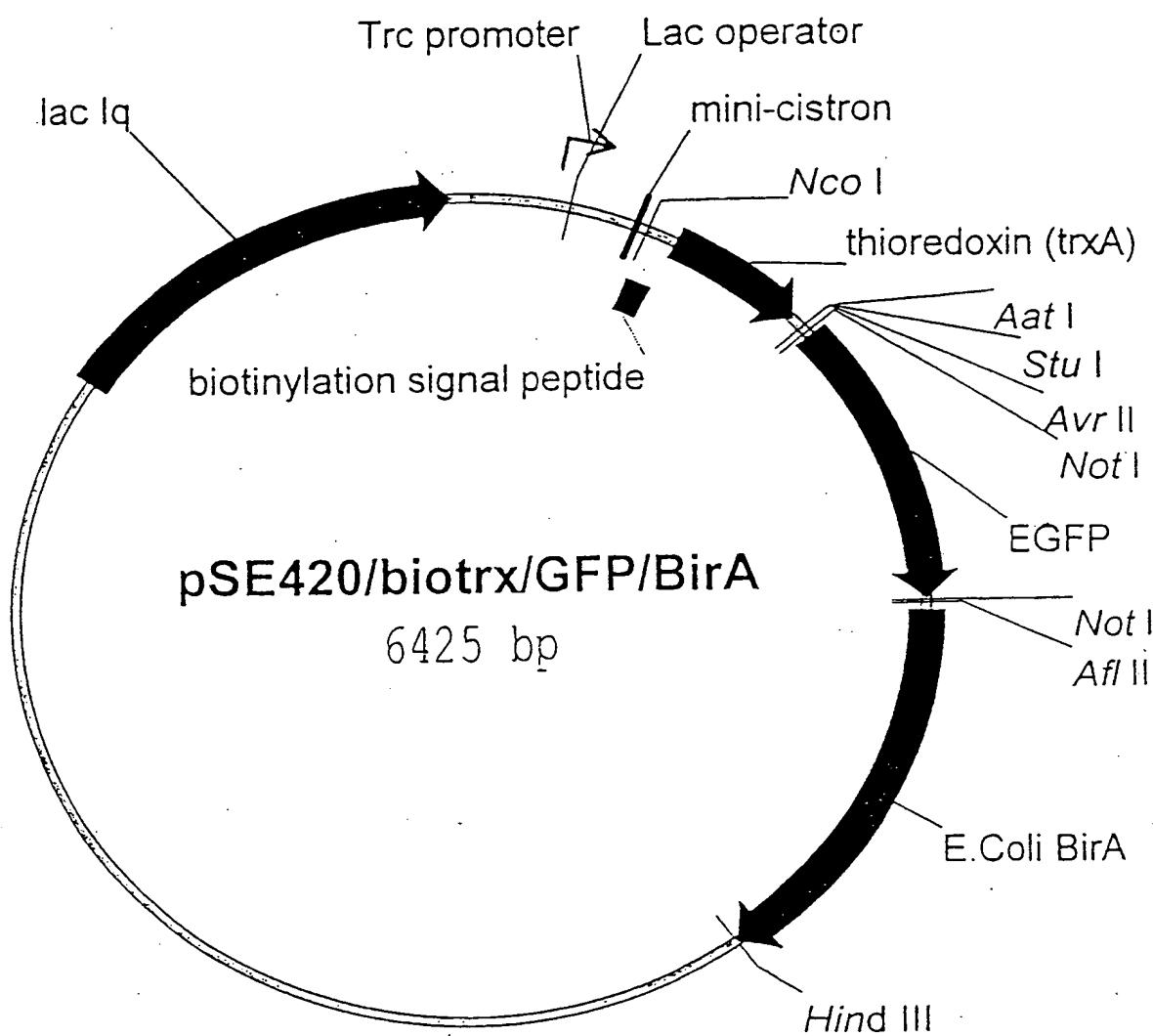


FIGURE 2

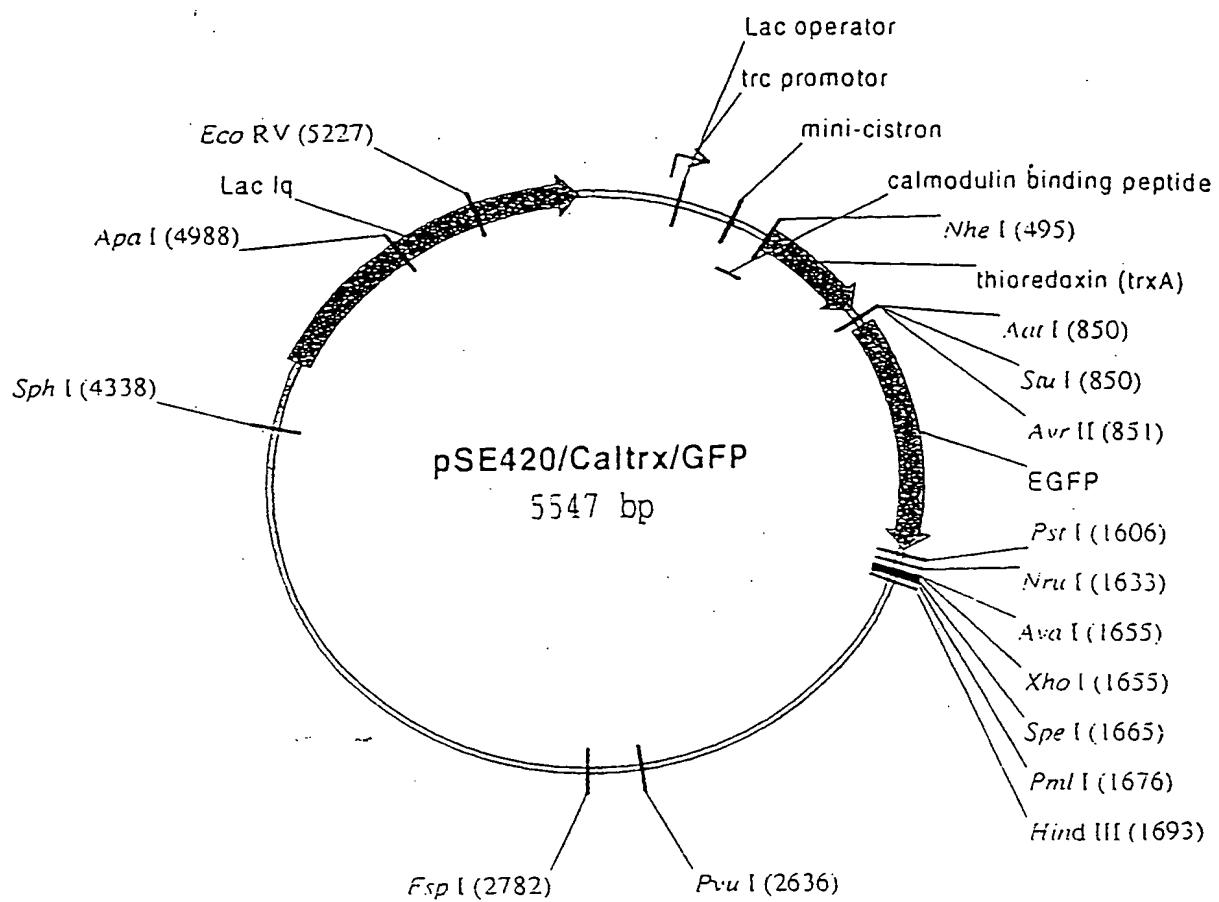


FIGURE 3

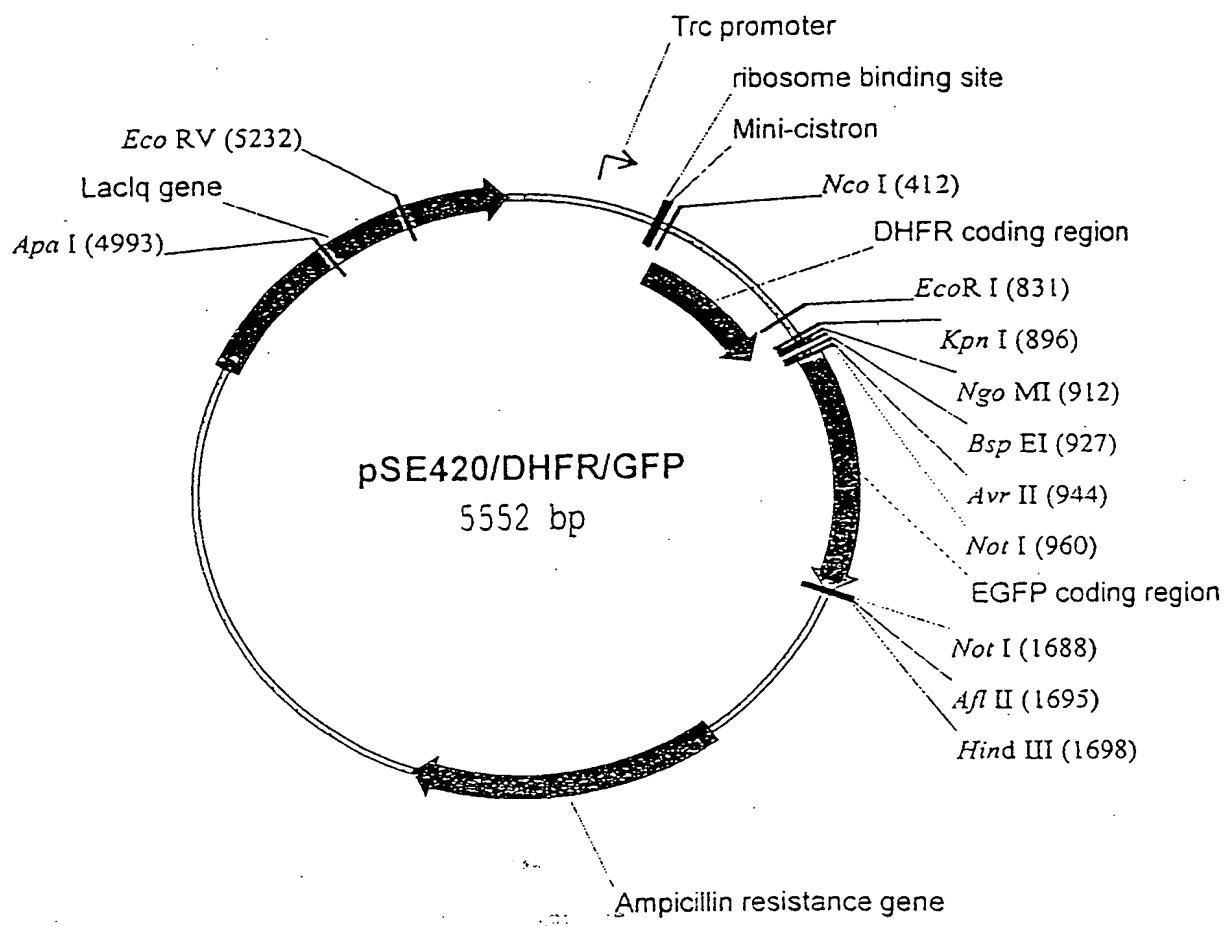


FIGURE 4

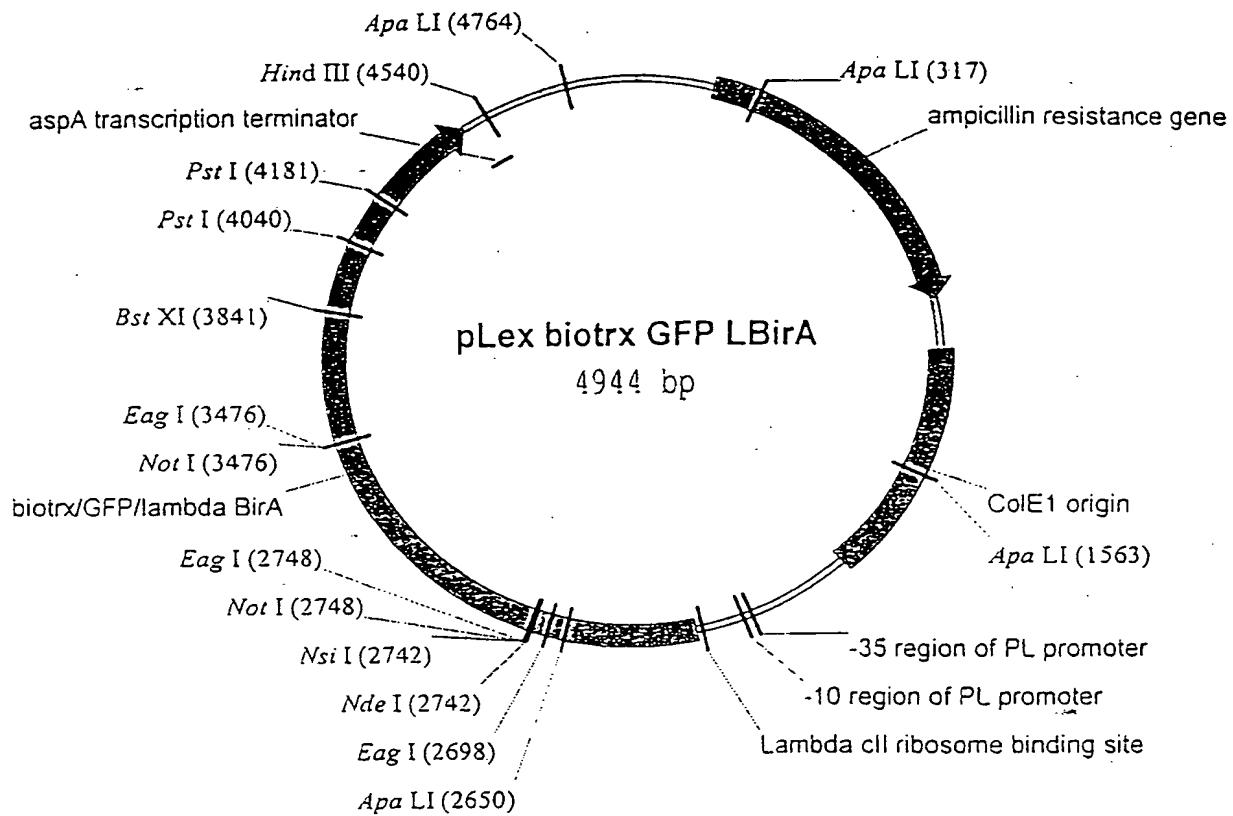


FIGURE 5

MTX-SANPAH bead schematic

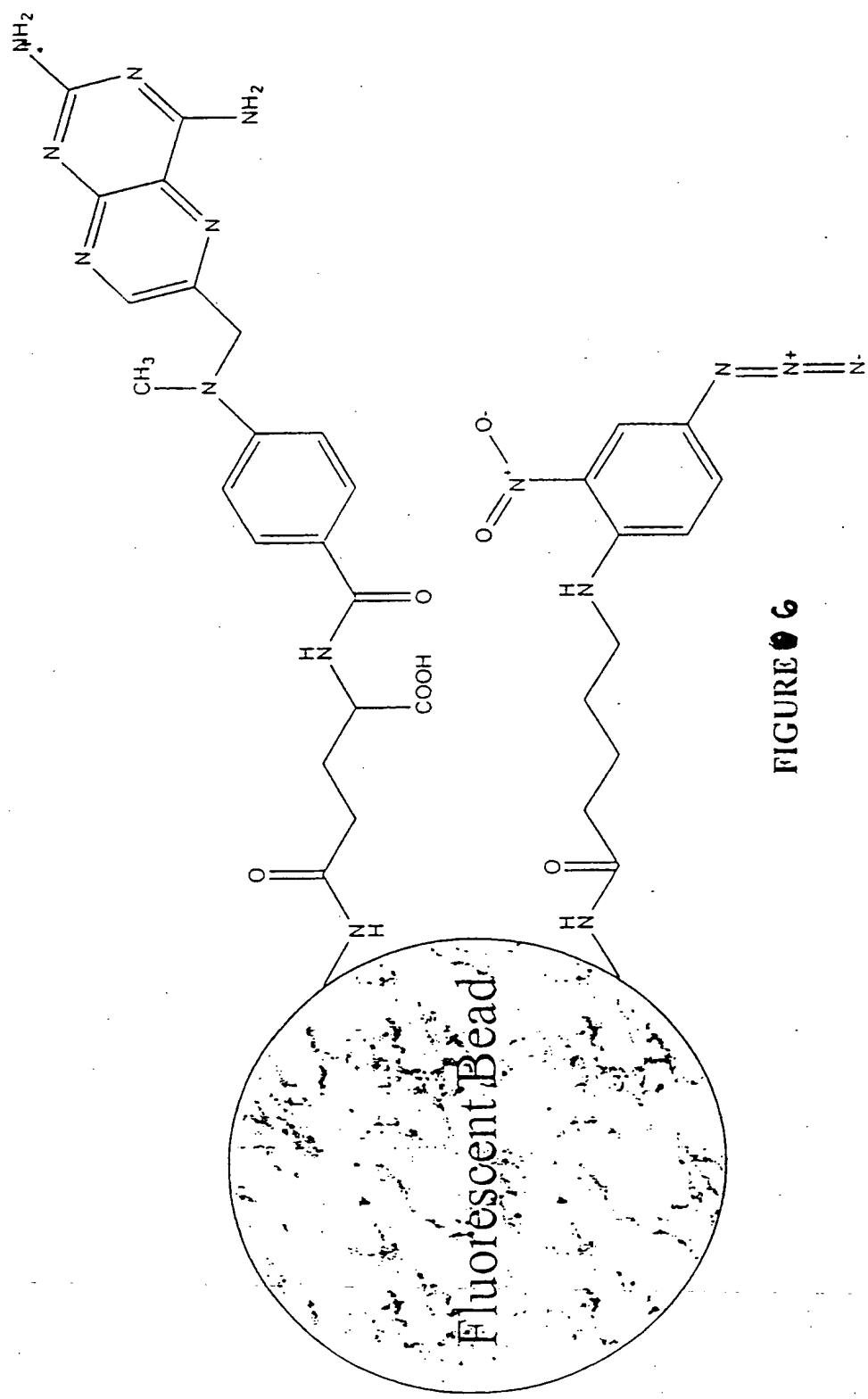
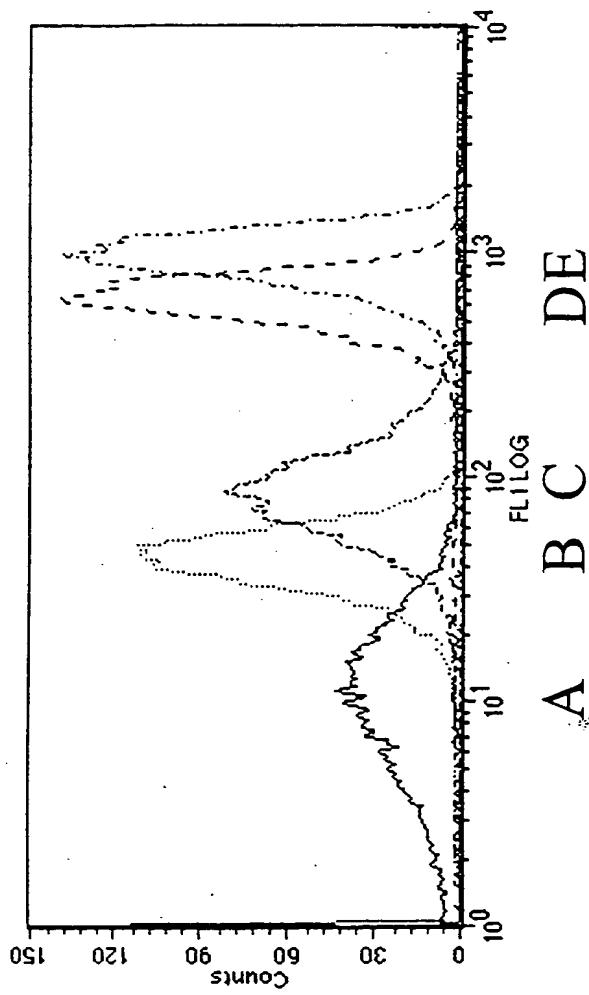


FIGURE 6



- A. Streptavidin coated particle reacted with BL21 lysate and FITC-calmodulin conjugate.
- B. Streptavidin coated particle reacted with a lysate having a biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate then EGTA.
- C. Streptavidin coated particle reacted with a lysate having a biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate.
- D. Streptavidin coated particle reacted with a lysate having a biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate and protein crosslinking agent.
- E. Streptavidin coated particle reacted with a lysate having a biotin-thioredoxin-CBP fusion protein, FITC-calmodulin conjugate, protein crosslinking agent then EGTA.

Figure 7.

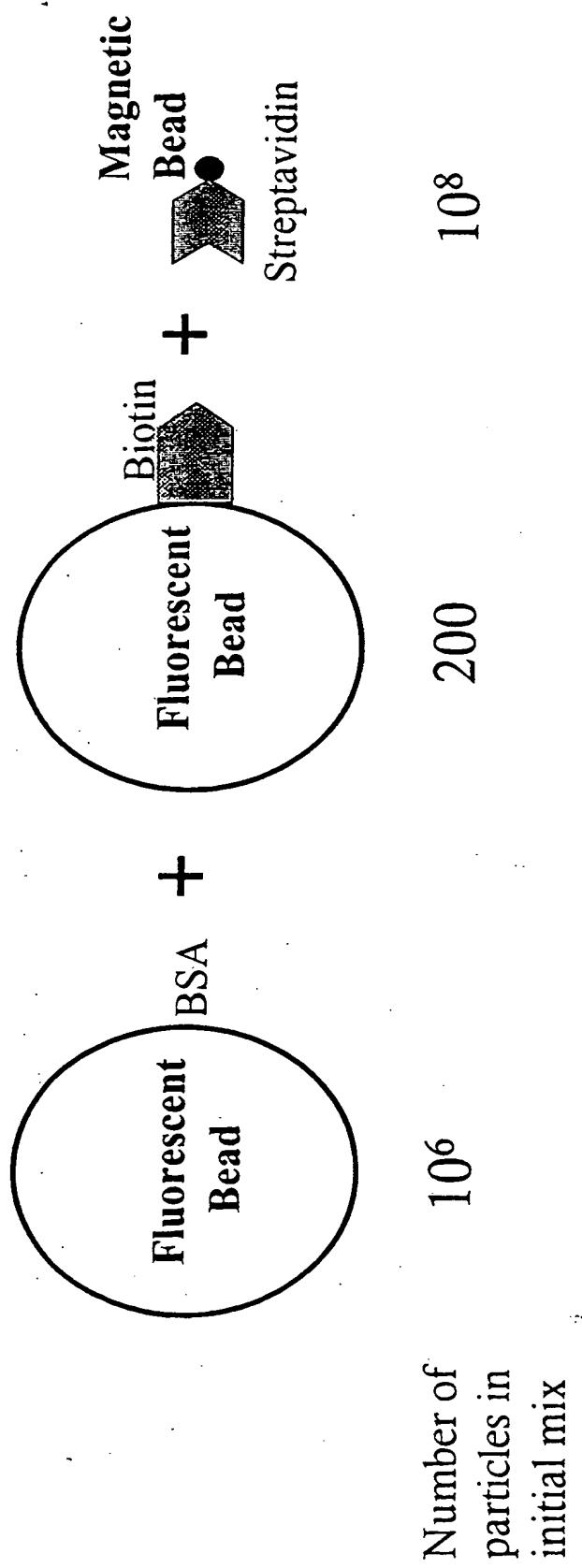


Figure 8.

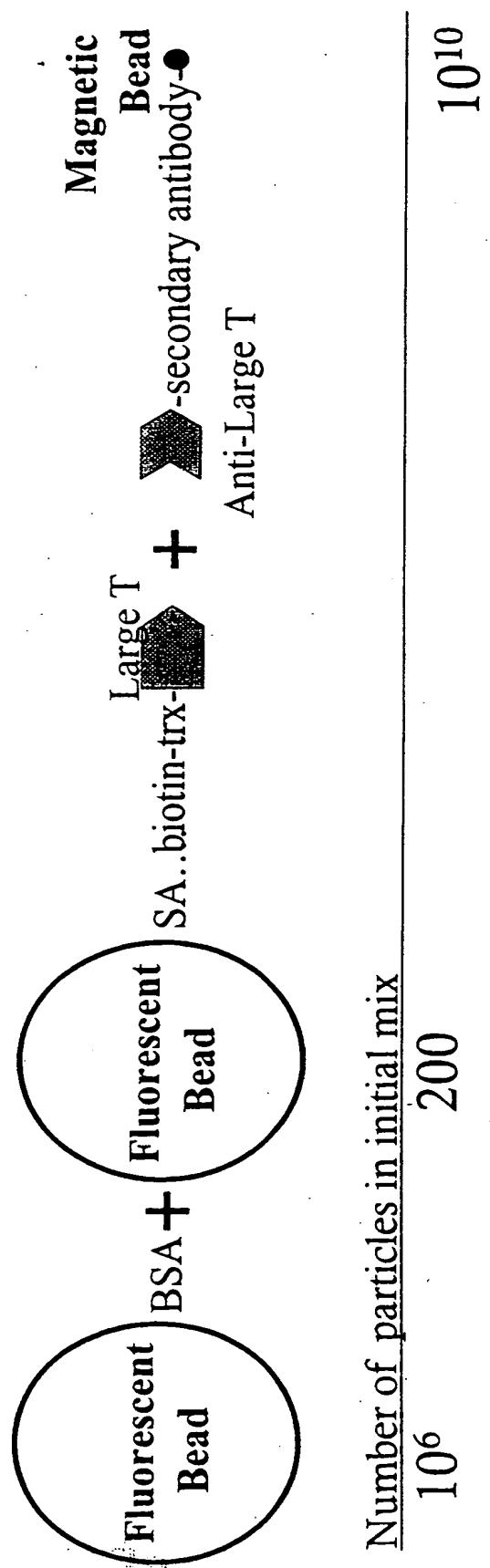


Figure 9.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/04089

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G01N C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98 26098 A (FELDHAUS MICHAEL JOHN ; KAMB ALEXANDER (US); VENTANA GENETICS INC () 18 June 1998 (1998-06-18) abstract claims ----	1-24
A	WO 98 53093 A (EBRIGHT RICHARD H ; SEUL MICHAEL (US); UNIV RUTGERS (US); BIOARRAY) 26 November 1998 (1998-11-26) abstract page 7, paragraph 3 -page 88, paragraph 3 ----	1-24
A	WO 95 16209 A (CIBA GEIGY AG ; FELDER EDUARD (CH); RINK HANS (CH); MATTHEWS IAN TI) 15 June 1995 (1995-06-15) claims ----	1-24
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

14 July 2000

Date of mailing of the international search report

26/07/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Pellegrini, P

INTERNATIONAL SEARCH REPORTInternational Application No
PCT/US 00/04089**C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	WO 99 49315 A (RIGEL PHARMACEUTICALS INC) 30 September 1999 (1999-09-30) claims -----	1-24

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 25

A very large number of possible products (human protein interaction maps) could fall within present claim 25. No support within the meaning of Article 6 PCT and/or disclosure within the meaning of article 5 PCT was found for such human protein interaction maps. In the present case, claim 25 so lacks support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Claim 25 also shows a lack of clarity within the meaning of Article 6 PCT such that a meaningful search of this claim was deemed impossible. Furthermore, claim 25 can not be searched by this International Searching Authority because it relates to mere presentation of information, within the meaning of Article 17(2)(b) and Rule 39(v) PCT.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/04089

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9826098	A 18-06-1998	US AU EP	6060240 A 5605798 A 0948649 A	09-05-2000 03-07-1998 13-10-1999
WO 9853093	A 26-11-1998	AU EP	7599698 A 1003904 A	11-12-1998 31-05-2000
WO 9516209	A 15-06-1995	AU CA EP FI HU JP NO	1068295 A 2175078 A 0733213 A 962371 A 75223 A 9508355 T 962186 A	27-06-1995 15-06-1995 25-09-1996 07-06-1996 28-04-1997 26-08-1997 29-05-1996
WO 9949315	A 30-09-1999	AU	3364699 A	18-10-1999

This Page Blank (uspto)